Kuolevuuden estimointiongelma ja Nelson-Aalen estimaattori henkivakuutusmatematiikassa

Marianne Närhi

Pro Gradu -tutkielma

12. huhtikuuta 2020

Matemaattis-luonnontieteellinen tiedekunta
Matematiikan ja tilastotieteen osasto
Helsingin yliopisto

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Koulutusohjelma — Utbildningsprogram — Degree programme
Matemaattis-luonnontieteellinen	Matematiikan ja tilastotieteen maisteriohjelma

Tekijä — Författare — Author

Marianne Närhi

Työn nimi — Arbetets titel — Title

Kuolevuuden estimointiongelma ja Nelson-Aalen estimaattori henkivakuutusmatematiikassa

Oppiaine — Läroämne — Subject

Matematiikka

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Pro gradu -tutkielma	Huhtikuu 2020	63 s.

Tiivistelmä — Referat — Abstract

Henkivakuutusyhtiöt tarjoavat asiakkailleen monenlaisia tuotteita. Vakuutuksia on erityyppisiä, mutta usein ne ovat liitoksissa vakuutetun elinaikaan. Mainittakoon näistä esimerkiksi kuolemanvara- ja elämänvaravakuutus. Ensimmäisessä korvaus maksetaan mikäli vakuutettu kuolee vakuutusaikana ja toisessa mikäli vakuutettu on elossa ennalta sovittuna ajanhetkenä. Vakuutetun elinaika ei kuitenkaan ole tiedossa sopimusta tehdessä, joten vakuutusyhtiön pitää pystyä estimoimaan vakuutettujen kuolevuutta. Riittävän tarkalla estimoinnilla pyritään estämään tilanne, jossa korvausten määrä ylittää vakuutusyhtiön varat. Kuolevuusennustetta voidaan käyttää muun muassa vakuutusten hinnoitteluun. Estimointi on kuitenkin haastavaa, sillä kuolevuuden kehitykseen tulevaisuudessa vaikuttavat muun muassa mahdolliset lääketieteelliset läpimurrot tai populaation elintapojen muutokset. Kuolevuus ei pysy samana sukupolvesta toiseen, vaan pääsääntöisesti monissa maissa uusi sukupolvi elää edellistä sukupolvea keskimäärin pidempään. Kuolevuutta onkin helpompi ennustaa lyhyellä kuin pitkällä aikavälillä.

Tutkielman alussa määrittelemme tämän työn kannalta oleellisia esitietoja, jotka liittyvät sekä elinaikaan ja kuolevuuteen että yleisesti stokastisiin prosesseihin. Erityisen tärkeitä ovat elinajan ja kuolevuusfunktion käsite. Näiden lisäksi martingaali, laskuriprosessi ja kompensaattori ovat tämän työn avainkäsitteitä. Tutustumme määritelmien lisäksi Doob-Meierin hajotelmaan, jonka perusteella alimartingaali voidaan kirjoittaa systemaattisen ja täysin satunnaisen osan summana. Systemaattisesta osasta puhutaan kompensaattorina ja satunnaisen osan muodostaa martingaali.

Tutkielman tarkoituksena on johtaa kumulatiivista kuolevuutta estimoiva Nelson-Aalen estimaattori tilanteessa, jossa vakuutettuja on n kappaletta ja vakuutetun mahdollisia eri kuolinsyitä k kappaletta. Oletamme parametrin n arvon olevan suhteellisen suuri ja parametrin k arvon suhteellisen pieni. Johdamme lisäksi estimaattorin odotusarvon sekä varianssin. Havaitaan, että estimaattori on hieman harhainen, mutta kuitenkin asymptoottisesti harhaton. Teemme lisäksi lyhyen sovelluksen R:llä, jonka tarkoituksena on auttaa lukijaa hahmottamaan miltä todellisen otoksen pohjalta laaditut Nelson-Aalen estimaatit voisivat näyttää ja tutkitaan kuinka hyvin ne vastaavat todellisia arvoja.

Tutkielman loppupuolella tarkastellaan tilannetta, jossa vakuutettujen määrä kasvaa rajatta ja huomataan, että normalisoitu Nelson-Aalen estimaattori alkaa muistuttaa Gaussista martingaalia. Erityisesti kiinteällä ajanhetkellä estimaattori on asymptoottisesti normaalijakautunut. Todistuksessa käytämme Rebolledon keskeistä raja-arvolausetta martingaaleille. Tulosta käyttämällä olisi mahdollista määrittää luottamusrajat estimoitavalle kumulatiiviselle kuolevuudelle. Lopuksi käymme läpi vaihtoehtoisia tapoja estimoida kuolevuutta.

Avainsanat — Nyckelord — Keywords

Nelson-Aalen estimaattori, kuolevuusfunktio, martingaali, laskuriprosessi, kompensaattori

Säilytyspaikka — Förvaringsställe — Where deposited

Kumpulan tiedekirjasto

Muita tietoja — Övriga uppgifter — Additional information

Sisältö

1 Johdanto

Henkivakuutusyhtiöt tarjoavat asiakkailleen monenlaisia tuotteita. Vakuutuksia on erityyppisiä, mutta usein ne ovat liitoksissa vakuutetun elinaikaan. Mainittakoon näistä esimerkiksi kuolemanvara- ja elämänvaravakuutus. Ensimmäisessä korvaus maksetaan mikäli vakuutettu kuolee vakuutusaikana ja toisessa mikäli vakuutettu on elossa ennalta sovittuna ajanhetkenä. Vakuutetun elinaika ei kuitenkaan ole tiedossa sopimusta tehdessä, joten vakuutusyhtiön pitää pystyä estimoimaan vakuutettujen kuolevuutta. Riittävän tarkalla estimoinnilla pyritään estämään tilanne, jossa korvausten määrä ylittää vakuutusyhtiön varat. Kuolevuusennustetta voidaan käyttää muun muassa vakuutusten hinnoitteluun. Estimointi on kuitenkin haastavaa, sillä kuolevuuden kehitykseen tulevaisuudessa vaikuttavat muun muassa mahdolliset lääketieteelliset läpimurrot tai populaation elintapojen muutokset. Kuolevuus ei pysy samana sukupolvesta toiseen, vaan pääsääntöisesti monissa maissa uusi sukupolvi elää edellistä sukupolvea keskimäärin pidempään. Kuolevuutta onkin helpompi ennustaa lyhyellä kuin pitkällä aikavälillä.

Erilaiset säännökset asettavat rajoituksia kuolevuusmalleille. Johdamme tässä työssä erään ei-parametrisen estimaattorin, Nelson-Aalen estimaattorin, ja tutkimme sen ominaisuuksia, kun vakuutettujen määrä kasvaa rajatta. Ei-parametrisella mallilla tarkoitetaan, että ongelman asettelussa ei oleteta olevan perhettä, jonka parametrit määräävät.

Aloitamme esittelemällä luvussa 2 tämän työn kannalta oleellisia esitietoja, jotka liittyvät sekä elinaikaan ja kuolevuuteen että yleisesti stokastisiin prosesseihin. Erityisen tärkeitä ovat elinajan ja kuolevuusfunktion käsite. Näiden lisäksi martingaali, laskuriprosessi ja kompensaattori ovat tämän työn avainkäsitteitä. Tutustumme määritelmien lisäksi Doob-Meierin hajotelmaan, jonka perusteella alimartingaali voidaan kirjoittaa systemaattisen ja täysin satunnaisen osan summana. Systemaattisesta osasta puhutaan kompensaattorina ja satunnaisen osan muodostaa martingaali. Luvun lopussa annetaan esimerkki erään laskuriprosessin kompensaattorista.

Luvussa 3 muodostetaan Nelson-Aalen estimaattori tilanteessa, jossa vakuutettuja on n kappaletta ja vakuutetun mahdollisia eri kuolinsyitä k kappaletta. Oletamme parametrin n arvon olevan suhteellisen suuri ja parametrin k arvon suhteellisen pieni. Lisäksi johdetaan estimaattorin odotusarvo sekä varianssi. Havaitaan, että estimaattori on hieman harhainen, mutta kuitenkin asymptoottisesti harhaton. Luvun lopussa tehdään sovellus R:llä, jonka tarkoituksena on auttaa lukijaa hahmottamaan miltä todellisen otoksen pohjalta laaditut Nelson-Aalen estimaatit voisivat näyttää ja tutkitaan kuinka hyvin ne vastaavat todellisia arvoja.

Luvussa 4 annetaan vakuutettujen määrän kasvaa rajatta ja osoitetaan, että tällöin estimaattori alkaa muistuttaa Gaussista martingaalia. Kiinteällä ajanhetkellä puhutaan Nelson-Aalen estimaattorin asymptoottisesta normaaliuudesta. Todistamme ensin lauseen käyttämällä Rebolledon keskeistä raja-arvolausetta martingaaleille, jonka jälkeen osoitamme, että lauseen ehdot pätevät luvun 3 kilpailevien kuolinsyiden tilanteessa. Tulosta käyttämällä on mahdollista määrittää luottamusrajat estimoitavalle kumulatiiviselle kuolevuudelle, mutta sivuutamme sen tässä työssä.

Luvussa 5 käydään läpi muita mahdollisia tapoja estimoida kuolevuutta. Tutustumme parametrisoiduista malleista muun muassa Weibullin, Makehamin ja Gompertzin malleihin. Esitellään lisäksi lyhyesti kuolevuustaulu sekä referenssikuolevuudet k2004 ja k2012. Lopuksi sanotaan muutama sana Kaplan-Meierin estimaattorista.

Luvussa 6 kootaan lyhyesti tehtyjä huomioita ja pohditaan hieman millaisissa tilanteissa mahdollisesti kannattaisi käyttää Nelson-Aalen estimaattoria.

Tutkielman päälähteenä on Per Kragh Andersenin, Ørnulf Borganin, Richard D. Gillin ja Niels Keidingin kirja *Statistical models based on counting processes* ja siitä erityisesti luku IV *Nonparametric estimation*.

2 Esitiedot

Oletetaan läpi tutkielman, että on olemassa todennäköisyysavaruus $(\Omega, \mathscr{F}, \mathbb{P})$, jossa tarkastelemme aikaväliä $[0, b]$, missä $b \in (0, \infty)$.

2.1 Elinaika ja kuolevuusfunktio

Tässä luvussa seurataan David C. M. Dicksonin, Mary R. Hardyn ja Howard R. Watersin kirjan *Actuarial mathematics for life contingent risks* [7] merkintä- ja esitystapaa. Merkitään vakuutetun ikää muutujalla x, jolle pätee $x \geq 0$.

2.1.1 Elinaika

Vakuutetun ja vakuutusyhtiön solmiessa henkivakuutussopimusta vakuutetun elinaika ei ole tiedossa. Vakuutetun elinaika oletetaan jatkuvasti jakautuneeksi ja merkitään sitä satunnaismuuttujalla $T(x)$, jossa x on vakuutetun ikä. Syntymähetkellä vakuutetun koko elinaikaa merkitään $T = T(0)$ ja vastaavasti x-ikäisen vakuutetun jäljellä olevaa elinaikaa $T(x)$. Tällöin x-ikäisen vakuutetun ikä kuolinhetkellä on $x + T(x)$.

Olkoon F_x satunnaismuuttujan $T(x)$ kertymäfunktio, toisin sanoen

$$F_x(t) = \mathbb{P}\Big(T(x) \leq t\Big),$$

missä $t \in \mathbb{R}$, jonka arvo on 0 kaikilla negatiivisilla parametrin t arvoilla. Kertymäfunktion F_x voi tulkita todennäköisyytenä, jolla x-ikäinen vakuutettu kuolee ennen ikää $x + t$.

Määritelmä 2.1. Olkoot F_x ja $T(x)$, kuten edellä.
Tällöin

$$S_x(t) = 1 - F_x(t) = \mathbb{P}\Big(T(x) > t\Big),$$

missä funktiota S_x kutsutaan selviytymisfunktioksi.

Päinvastoin kuin elinajan kertymäfunktio, selvitymisfunktio kuvaa todennäköisyyttä, että x-ikäinen vakuutettu elää vielä ainakin t vuotta. Kertymäfunktion perusominaisuuksista ja elinajan jakauman jatkuvuudesta johtuen selvitymisfunktio toteuttaa seuraavat ehdot: $S_x(0) = 1$, $\lim_{t\to\infty} S_x(t) = 0$ ja $S_x(t)$ on vähenevä muuttujan t funktio.

Huomautus 2.1. Käytetään funktioista F_0 ja S_0 lyhennysmerkintöjä F ja S.

Aiempien määritelmien nojalla kaikilla $x \geq 0$ ja $t > 0$ pätee

$$\mathbb{P}\Big(T(x) \leq t\Big) = \mathbb{P}\Big(T(0) \leq x + t \,|\, T(0) > x\Big). \tag{2.1}$$

Tällöin käyttämällä ehdollisen todennäköisyyden määritelmää saadaan

$$\mathbb{P}\Big(T(x) \le t\Big) = \frac{\mathbb{P}\Big(x < T(0) \le x+t\Big)}{\mathbb{P}\Big(T(0) > x\Big)} = \frac{F_0(x+t) - F_0(x)}{1 - \mathbb{P}\Big(T(0) \le x\Big)} = \frac{F_0(x+t) - F_0(x)}{S_0(x)}.$$

Tätä tulosta käyttämällä saadaan selvitymisfunktio muotoon

$$\begin{aligned} S_x(t) &= 1 - \mathbb{P}\Big(T(x) \le t\Big) \\ &= 1 - \frac{F_0(x+t) - F_0(x)}{S_0(x)} \\ &= \frac{S_0(x) - \Big(1 - S_0(x+t)\Big) + 1 - S_0(x)}{S_0(x)} \\ &= \frac{S_0(x+t)}{S_0(x)}. \end{aligned}$$

Saamme siis esityksen

$$S_x(t) = \frac{S_0(x+t)}{S_0(x)}. \tag{2.2}$$

Selvitymis- ja kuolintodennäköisyyksistä käytetään usein aktuaarisia lyhennysmerkintöjä, joita esiintyy paljon kirjallisuudessa. Esitellään ne seuraavaksi.

Määritelmä 2.2. Standardit aktuaariset notaatiot selviytymis- ja kuolintodennäköisyyksille ovat

$${}_tp_x = \mathbb{P}\Big(T(x) > t\Big) = S_x(t) \qquad \text{ja} \qquad {}_tq_x = \mathbb{P}\Big(T(x) \le t\Big) = 1 - S_x(t) = F_x(t),$$

missä ${}_tp_x$ tulkitaan todennäköisyytenä, että x-ikäinen vakuutettu selviää ainakin ikään $x+t$ ja ${}_tq_x$ todennäköisyytenä, että x-ikäinen vakuutettu kuolee ennen ikää $x+t$.

Huomautus 2.2. Notaatioista ${}_1p_x$ ja ${}_1q_x$ käytetään usein lyhennysmerkintöjä p_x ja q_x.

2.1.2 Kuolevuusfunktio

Määritelmä 2.3 (Kuolevuusfunktio)**.** Merkitään $\mu(x)$ vastasyntyneen vakuutetun kuolevuutta ajanhetkellä x ja määritellään funktio $\mu : [0,\infty) \to [0,\infty)$ siten, että

$$\mu(x) = \lim_{y \to 0^+} \frac{1}{y} \mathbb{P}\Big(T(0) \le x + y \,|\, T(0) > x\Big).$$

Käyttämällä hyödyksi esitystapaa (2.1) kuolevuus voidaan kirjoittaa myös muodoissa

$$\mu(x) = \lim_{y \to 0^+} \frac{1}{y} \mathbb{P}\Big(T(x) \leq y\Big) = \lim_{y \to 0^+} \frac{1}{y} \Big(1 - S_x(y)\Big) \tag{2.3}$$

Intuitiivisen ymmärryksen kuolevuudesta saa approksimaatiosta

$$\mu(x)\mathrm{d}x \approx \mathbb{P}\Big(T(0) \leq x + \mathrm{d}x \,|\, T(0) > x\Big),$$

jonka voi tulkita todennäköisyytenä, että x-ikäinen vakuutettu kuolee ennen ikää $x + \mathrm{d}x$, missä $\mathrm{d}x > 0$ on hyvin pieni.

Huomautus 2.3. Muualla kirjallisuudessa käsitelty *failure rate* tai *hazard rate* määritellään usein matemaattisesti samalla tavalla kuin kuolevuus. Niissä tehty tarkastelu sopii täten myös kuolevuusfunktioon.

Määritelmä 2.4. Määritellään elinaikaan $T(x)$ liittyvä kuolevuus

$$\mu^x(t) = \mu(x + t)$$

kaikille $x \geq 0$ ja $t \geq 0$.

Osoitetaan seuraavassa lauseessa kuolevuusfunktion yhteys notaatioihin ${}_tp_x$ ja ${}_tq_x$.

Lause 2.1. *Oletetaan, että $S_x(t)$ on derivoituva kaikilla $t > 0$. Oletetaan lisäksi, että $x \geq 0$ ja $t > 0$. Tällöin*

$$ {}_tp_x = \exp\left(-\int_x^{x+t} \mu(s)\,\mathrm{d}s\right) = \exp\left(-\int_0^t \mu(x+s)\,\mathrm{d}s\right) \qquad \textit{ja}$$

$$ {}_tq_x = 1 - \exp\left(-\int_x^{x+t} \mu(s)\,\mathrm{d}s\right) = 1 - \exp\left(-\int_0^t \mu(x+s)\,\mathrm{d}s\right).$$

Lauseen 2.1. todistus. Voimme kirjoittaa lausekkeen (2.2) muodossa

$$S_x(y) = \frac{S_0(x+y)}{S_0(x)}.$$

Käyttämällä tätä muotoa selviytymisfunktiosta kirjoitetaan kuolevuus $\mu(x)$ kuten lausekkessa (2.3)

$$\begin{aligned} \mu(x) &= \lim_{y \to 0^+} \frac{1}{y}\Big(1 - S_x(y)\Big) \\ &= \lim_{y \to 0^+} \frac{1}{y}\left(1 - \frac{S_0(x+y)}{S_0(x)}\right) \\ &= \frac{1}{S_0(x)} \lim_{y \to 0^+} \frac{S_0(x) - S_0(x+y)}{y} \\ &= -\frac{1}{S_0(x)}\left(\frac{\mathrm{d}}{\mathrm{d}x} S_0(x)\right) \end{aligned} \tag{2.4}$$

Viimeinen yhtäsuuruus saadaan käyttämällä derivaatan määritelmää. Tiedämme, että funktiolle h pätee

$$\frac{\mathrm{d}}{\mathrm{d}x}\log h(x) = \frac{1}{h(x)}\frac{\mathrm{d}}{\mathrm{d}x}h(x),$$

mikäli derivaatta on olemassa. Täten saamme lausekkeen (2.4) muotoon

$$\mu(x) = -\frac{\mathrm{d}}{\mathrm{d}x}\log S_0(x).$$

Kun tätä integroidaan yli välin $(0, y)$, saadaan

$$\int_0^y \mu(x)\,\mathrm{d}x = \int_0^y -\frac{\mathrm{d}}{\mathrm{d}x}\log S_0(x)\,\mathrm{d}x = -\Big(\log S_0(y) - \log S_0(0)\Big).$$

Nyt $\log S_0(0) = \log \mathbb{P}(T_0 > 0) = \log 1 = 0$, joten saadaan

$$\int_0^y \mu(x)\,\mathrm{d}x = -\log S_0(y). \tag{2.5}$$

Ratkaisemalla yhtälö (2.5) termin $S_0(y)$ suhteen päädytään yhtälöön

$$S_0(y) = \exp\left(-\int_0^y \mu(x)\,\mathrm{d}x\right).$$

Lopulta kirjoittamalla selviytymisfunktio muodossa (2.2) saadaan

$$\begin{aligned}
{}_tp_x = S_x(t) &= \frac{S_0(x+t)}{S_0(x)} \\
&= \frac{\exp\Big(-\int_0^{x+t}\mu(r)\,\mathrm{d}r\Big)}{\exp\Big(-\int_0^{x}\mu(r)\,\mathrm{d}r\Big)} \\
&= \exp\left(-\int_0^{x+t}\mu(r)\,\mathrm{d}r + \int_0^x \mu(r)\,\mathrm{d}r\right) \\
&= \exp\left(-\int_x^{x+t}\mu(r)\,\mathrm{d}r\right) \\
&= \exp\left(-\int_0^{t}\mu(x+r)\,\mathrm{d}r\right)
\end{aligned}$$

Todistus notaatiolle ${}_tq_x$ saadaan tehtyä käyttämällä määritelmää ${}_tq_x = 1 - S_x(t)$ ja edellistä todistusta. □

Esitellään vielä yksi esitystapa kuolevuusfunktiolle. Merkitään sitä varten kertymäfunktioon F_x liittyvää tiheysfunktiota f_x. Tiedämme, että

$$f_x(t) = \frac{\mathrm{d}}{\mathrm{d}t}F_x(t) = \frac{\mathrm{d}}{\mathrm{d}t}\Big(1 - S_x(t)\Big) = -\frac{\mathrm{d}}{\mathrm{d}t}S_x(t).$$

Erityisesti pätee

$$f_0(t) = -\frac{\mathrm{d}}{\mathrm{d}t} S_0(t).$$

Täten voimme kirjoittaa lausekkeen (2.4) muotoon

$$\mu(x) = -\frac{1}{S_0(x)} \left(\frac{\mathrm{d}}{\mathrm{d}x} S_0(x) \right) = \frac{f_0(x)}{S_0(x)} = \frac{f_0(x)}{1 - F_0(x)}.$$

Esitellään seuraavaksi lyhyitä esimerkkejä kuolevuusfunktiosta.

Esimerkki 2.1. Olkoon T_0 eksponenttijakautunut parametrilla $\lambda > 0$. Tällöin

$$F_0(x) = 1 - e^{-\lambda x} \qquad \text{ja} \qquad f_0(x) = \lambda e^{-\lambda x}.$$

Kuolevuusfunktio on tässä tapauksessa vakio

$$\mu(x) = \frac{f_0(x)}{1 - F_0(x)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda.$$

Tämä esimerkki ei todellisuudessa ole kovin realistinen, sillä esimerkissä kuolevuus pysyy samana iästä riippumatta.

Esimerkki 2.2. Oletetaan, että T_0 on jakautunut siten, että

$$F_0(x) = 1 - e^{-x^2} \qquad \text{ja} \qquad f_0(x) = 2xe^{-x^2},$$

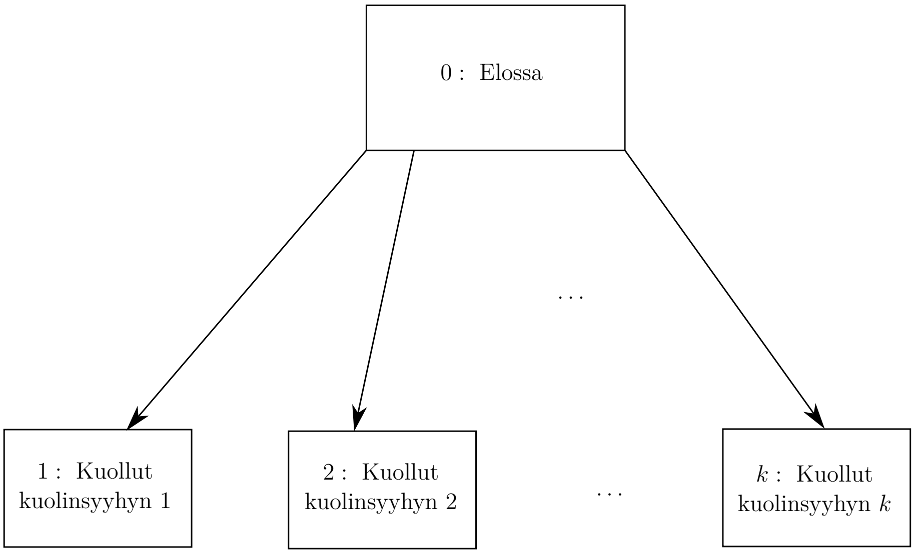
missä $x \geq 0$. Kuolevuusfunktioksi saadaan

$$\mu(x) = \frac{2xe^{-x^2}}{1 - (1 - e^{-x^2})} = 2x.$$

Nyt malli on jo hieman realistisempi, sillä kuolevuus kasvaa iän myötä. Kasvu on kuitenkin lineaarista, mikä ei välttämättä ole totta oikeassa elämässä.

2.1.3 Kilpailevista kuolinsyistä

Olemme tässä työssä erityisen kiinnostuneita niin sanotusta kilpailevien kuolinsyiden teorista. Oletetaan, että tarkastelemme yhtä vakuutettua, jonka mahdollisia eri kuolinsyitä on k kappaletta.



Kuva 2.1: Asetelma, jossa vakuutetulla on mahdollisia eri kuolinsyitä k kappaletta. Mahdollisia siirtymiä tilasta toiseen kuvataan nuolilla.

Merkitään jokaiseen kuolinsyyhyn liittyvää vakuutetun hypoteettista elinaikaa satunnaismuuttujalla T_h, missä $h \in \{1, ..., k\}$. Satunnaismuuttuja T_h kuvaa vakuutetun elinaikaa tilanteessa, jossa olisi olemassa vain kuolinsyy h. Ainoastaan yksi näistä elinajoista havaitaan ja se on pienin kuolinsyihin liittyvistä elinajoista. Vakuutetun todellinen elinaika T voidaan siis kirjoittaa muodossa

$$T = \min(T_1, ..., T_k).$$

Vakuutusyhtiö havaitsee vakuutetun kuollessa vain parin (T, H), missä H, joka saa arvoja joukosta $\{1, ..., k\}$, kertoo vakuutetun kuolinsyyn. Merkitään elinaikaan T_h liittyvää kuolevuutta μ_h.

Tutkitaan sitten tilannetta, jossa ajanhetkellä 0 vakuutettu on 0-vuotiaan sijaan x-vuotias. Merkitään vastaavasti $T(x) = \min(T_1(x), ..., T_k(x))$ ja jäljellä olevaan elinaikaan $T_h(x)$ liittyvää kuolevuutta μ_h^x. Merkitään lisäksi todennäköisyyttä, että ajanhetkellä 0 x-ikäinen vakuutettu on kuollut kuolinsyyhyn h ennen ajanhetkeä t seuraavasti:

$$G_h(t) = \mathbb{P}\Big(T(x) \leq t, \text{ kuolinsyy on } h\Big).$$

Esitellään seuraavaksi lause, joka kertoo kuolinsyyhyn h liittyvän kuolevuuden μ_h^x ja G_h yhteyden.

Lause 2.2. *Oletetaan, että satunnaismuuttujat $T_1(x), ..., T_k(x)$ ovat toisistaan riippumattomia.*

Tällöin

$$\mu_h^x(t) = \frac{G_h'(t)}{1 - F_x(t)}, \qquad \textit{missä } h = 1, ..., k$$

pätee mille tahansa $t > 0$, jolle pätee, että $F_x(t) < 1$ ja $G_h'(t)$ on olemassa.

Lauseen 2.2. todistus. Oletetaan, että $t > 0$ ja oletetaan, että $\mathbb{P}(T(x) > t) \in (0, 1)$. Olkoon $\Delta > 0$ pieni. Tällöin pätee

$$\begin{aligned}
G_h(t+\Delta) - G_h(t) &= \mathbb{P}\Big(T_h(x) \in (t, t+\Delta], T_r(x) > T_h(x) \text{ kaikilla } r \neq h\Big) \\
&\leq \mathbb{P}\Big(T_h(x) \in (t, t+\Delta], T_r(x) > t \text{ kaikilla } r \neq h\Big) \\
&\overset{\perp\!\!\!\perp}{=} \mathbb{P}\Big(T_h(x) \in (t, t+\Delta]\Big) \prod_{r \neq h} \mathbb{P}\Big(T_r(x) > t\Big) \\
&= \mathbb{P}\Big(T_h(x) \leq t + \Delta \,|\, T_h(x) > t\Big)\mathbb{P}\Big(T(x) > t\Big),
\end{aligned}$$

missä toiseksi viimeinen yhtäsuuruus seuraa hypotettisten elinaikojen riippumattomuudesta.

Toisaalta

$$\begin{aligned}
G_h(t+\Delta) - G_h(t) &= \mathbb{P}\Big(T_h(x) \in (t, t+\Delta], T_r(x) > T_h(x) \text{ kaikilla } r \neq h\Big) \\
&\geq \mathbb{P}\Big(T_h(x) \in (t, t+\Delta], T_r(x) > t + \Delta \text{ kaikilla } r \neq h\Big) \\
&\overset{\perp\!\!\!\perp}{=} \mathbb{P}\Big(T_h(x) \in (t, t+\Delta]\Big) \prod_{r \neq h} \mathbb{P}\Big(T_r(x) > t + \Delta\Big) \\
&= \mathbb{P}\Big(T_h(x) \leq t + \Delta \,|\, T_h(x) > t\Big) \frac{\mathbb{P}\Big(T_h(x) > t\Big)}{\mathbb{P}\Big(T_h(x) > t + \Delta\Big)} \mathbb{P}\Big(T(x) > t + \Delta\Big).
\end{aligned}$$

Täten, kun $\Delta \to 0+$

$$G_h(t+\Delta) - G_h(t) = \mathbb{P}\Big(T_h(x) \leq t + \Delta \,|\, T_h(x) > t\Big)\Big(\mathbb{P}\Big(T(x) > t\Big) + o(1)\Big),$$

missä $o(1) \to 0$, kun $\Delta \to 0+$.

Nyt voidaan kirjoittaa

$$
\begin{aligned}
\mathbb{P}\Big(T_h(x) \le t + \Delta \,|\, T_h(x) > t\Big) &= \frac{\mathbb{P}\Big(t < T_h(x) \le t + \Delta\Big)}{\mathbb{P}\Big(T_h(x) > t\Big)} \\
&= \frac{\mathbb{P}\Big(T_h(x) \le t + \Delta\Big) - \mathbb{P}\Big(T_h(x) \le t\Big)}{\mathbb{P}\Big(T_h(x) > t\Big)} \\
&= \frac{1 - \exp\left(- \int_x^{x+t+\Delta} \mu_h(s)\, \mathrm{d}s\right) - 1 + \exp\left(- \int_x^{x+t} \mu_h(s)\, \mathrm{d}s\right)}{\exp\left(- \int_x^{x+t} \mu_h(s)\, \mathrm{d}s\right)} \\
&= - \exp\left(- \int_{x+t}^{x+t+\Delta} \mu_h(s)\, \mathrm{d}s\right) + 1 \\
&= 1 - \exp\left(- \int_t^{t+\Delta} \mu_h(x+s)\, \mathrm{d}s\right) \\
&= 1 - \exp\left(- \int_t^{t+\Delta} \mu_h^x(s)\, \mathrm{d}s\right).
\end{aligned}
$$

Koska

$$
e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + ... = 1 + x\left(1 + \frac{x}{2} + ...\right) = 1 + x\Big(1 + o(1)\Big),
$$

kun $x \to 0$, saadaan funktion μ_h^x jatkuvuuskohdissa t

$$
\begin{aligned}
\mathbb{P}\Big(T_h(x) \le t + \Delta \,|\, T_h(x) > t\Big) &= 1 - 1 + \int_t^{t+\Delta} \mu_h^x(s)\, \mathrm{d}s\,(1 + o(1)) \\
&= \mu_h^x(t)\Delta\Big(1 + o(1)\Big),
\end{aligned}
$$

kun $\Delta \to 0+$. Täten

$$
G_h'(t) = \mu_h^x(t)\Big(1 - F_x(t)\Big).
$$

Saamme oikeastaan vaan oikean derivaatan, sillä lähestymme termiä Δ vain positiiviselta puolelta, mutta koska G_h' on jatkuva, saamme derivaatan. □

Mikäli vakuutettuja on yhden sijaan n kappaletta, merkitään ajanhetkellä 0 x-ikäisen vakuutetun i todellista jäljellä olevaa elinaikaa $T^i(x)$ ja kuolinsyyhyn h liittyvää elinaikaa $T_h^i(x)$. Oletetaan parametrin n arvon olevan suhteellisen suuri ja parametrin k arvon suhteellisen pieni. Oletetaan, että vakuutetut ovat samankaltaisia ja -ikäisiä. Oletetaan myös vakuutettujen jäljellä olevat elinajat toisistaan riippumattomiksi. Lisäksi oletetaan elinajat eri kuolinsyiden suhteen samoin jakautuneiksi ja että vain vain yksi vakuutettu voi kuolla kerrallaan. Oletamme vielä, että yksikäsitteinen kuolinsyy on aina tiedossa.

Huomautus 2.4. Kun jatkossa määritellään jäljellä olevasta elinajasta $T^i(x)$ tai $T(x)$ riippuvia prosesseja, jätetään yksinkertaistamisen vuoksi x merkitsemättä prosessin parametrilistauksessa.

Lähteenä tässä kappaleessa on käytetty luentomonisteen Nyrhinen [14] lukua 3.

2.2 Todennäköisyysteoriasta

Nostetaan tässä luvussa esille todennäköisyysteoriasta muutamia aiheita Nelson-Aalen estimaattorin konstruoimista varten. Esitellään pääasiassa tarvittavia määritelmiä, mutta myös muutamia tuloksia sekä luvun lopussa konkreettinen esimerkki. Pääasiallisina lähteinä on käytetty Anderden et al. [3] lukua II, Lehtomaa [11], Izyurov [9] ja Billingsley [6].

Määritelmä 2.5 (Filtraatio)**.** Olkoon $(\Omega, \mathscr{F}, \mathbb{P})$ todennäköisyysavaruus. Alisigma-algebrojen perhettä

$$(\mathscr{F}_t : t \in [0, b]),$$

jolle pätee

$$\mathscr{F}_s \subseteq \mathscr{F}_t \subseteq \mathscr{F} \qquad \text{kaikilla } s < t$$

ja

$$\mathscr{F}_s = \cap_{t>s} \mathscr{F}_t \qquad \text{kaikilla } s \text{ (oikealta jatkuvuus)},$$

kutsutaan filtraatioksi tai historiaksi.

Intuitiivisesti $\mathscr{F}_t$ on ajanhetken t tieto. Filtraation $\mathscr{F}_t$ voi tulkita sisältävän kaikki tapahtumat, joista tiedetään ajanhetkeen t mennessä tapahtuuko tapahtuma.

Määritelmä 2.6 (Stokastinen prosessi)**.** Stokastinen prosessi X on joukko aika-indeksoituja satunnaismuuttujia $(X(t) : t \in [0, b])$ eli

$$X = \{X(t) : t \in [0, b]\}.$$

Usein myös kirjoitetaan tarkemmin $X(t, \omega)$, missä $t \in [0, b]$ ja $\omega \in \Omega$.

Määritelmä 2.7 (Cadlag)**.** Stokastista prosessia X sanotaan cadlag (continu à droite, limité à gauche) prosessiksi, jos sen polut $(X(t, \omega) : t \in [0, b])$ ovat oikealta jatkuvia sekä on olemassa vasemmat raja-arvot

$$X(t-, \omega) = \lim_{s \to t-} X(s, \omega) \qquad \text{kaikilla } t \in (0, b].$$

Määritelmä 2.8 (Sopiva prosessi)**.** Stokastista prosessia $X = \{X(t) : t \in [0, b]\}$ kutsutaan sopivaksi, mikäli $X(t)$ on $\mathscr{F}_t$ -mitallinen kaikilla t.

Määritelmä 2.9 (Ennustettava prosessi)**.** Stokastinen prosessi X on ennustettava, jos $X(t)$ on $\mathscr{F}_{t-}$-mitallinen kaikilla $t \in [0, b]$. Tässä $\mathscr{F}_{t-}$ on pienin sigma-algebra, joka sisältää kaikki filtraatiot $\mathscr{F}_s$, kun $s < t$.

Määritelmä 2.10 (Pysäytyshetki)**.** Olkoon τ on satunnaismuuttuja, joka saa arvoja joukosta $[0, b] \cup \{\infty\}$. Satunnaismuuttujaa τ kutsutaan pysäytyshetkeksi, jos kaikilla $t \in [0, b]$ pätee

$$\{\tau \leq t\} \in \mathscr{F}_t.$$

Pysäytyshetken määritelmää voi tuntua vaikealta ymmärtää, mutta intuitiivisesti voidaan ajatella, että meillä on joku tapahtuma tulevaisuudessa, jonka ajankohtaa ei tiedetä. Satunnaismuuttuja, joka kuvaa tapahtuman ajankohtaa on pysäytyshetki, jos jokaisella ajanhetkeillä tiedetään, onko tapahtuma tapahtunut.

Olkoon X stokastinen prosessi ja τ pysäytyshetki. Tällöin prosessia X^τ, joka määritellään

$$X^\tau(t) = X\Big(\min(t, \tau)\Big),$$

kutsutaan pysäytetyksi prosessiksi.

Määritelmä 2.11 (Lokalisointi)**.** Oletetaan, että on olemassa kasvava jono pysäytyshetkiä τ_n, jolle pätee

$$\mathbb{P}(\tau_n \geq t) \to 1, \qquad \text{kun } n \to \infty, \text{ kaikilla } t \in [0, b].$$

Sanotaan, että tietty ominaisuus pätee prosessille X lokaalisti, jos ominaisuus pätee prosessille $\mathbb{1}(\tau_n > 0) X^{\tau_n}$.

Määritellään seuraavaksi melkein varma suppeneminen, stokastinen suppeneminen, suppeneminen L^p mielessä sekä jakaumasuppeneminen. Muotoillaan lisäksi vahva suurten lukujen laki.

Määritelmä 2.12 (Melkein varma suppeneminen)**.** Olkoot $X, X_1, X_2, ...$ satunnaismuuttujia samassa todennäköisyysavaruudessa Ω. Sanotaan, että X_i suppenee melkein varmasti kohti satunnaismuuttujaa X, ja merkitään $X_i \overset{m.v.}{\to} X$, jos joukossa, jonka todennäköisyys on 1, pätee

$$X_i(\omega) \overset{i \to \infty}{\longrightarrow} X(\omega) \qquad \text{kaikilla } \omega.$$

Määritelmä 2.13 (Stokastinen suppeneminen)**.** Olkoot $X, X_1, X_2, ...$ satunnaismuuttujia samassa todennäköisyysavaruudessa Ω. Sanotaan, että X_i suppenee stokastisesti kohti satunnaismuuttujaa X, ja merkitään $X_i \overset{\mathbb{P}}{\to} X$, jos kaikille $\epsilon > 0$ pätee

$$\mathbb{P}(|X_i - X| > \epsilon) \overset{i \to \infty}{\longrightarrow} 0.$$

Määritelmä 2.14 (L^p suppeneminen)**.** Olkoon $p \geq 1$. Olkoot $X, X_1, X_2, ...$ satunnaismuuttujia samassa todennäköisyysavaruudessa Ω. Sanotaan, että X_i suppenee L^p:ssä kohti satunnaismuuttujaa X, ja merkitään $X_i \xrightarrow{L^p} X$, jos pätee

$$\mathbb{E}(|X_i - X|^p) \xrightarrow{i\to\infty} 0.$$

Määritelmä 2.15 (Jakaumasuppeneminen)**.** Olkoon $X_1, X_2, ...$ jono satunnaismuuttujia, joilla on kertymäfunktiot $F_1, F_2, ...$ Olkoon X satunnaismuuttuja, jolla on kertymäfunktio G.

Sanotaan, että jono X_i suppenee jakaumaltaan kohti satunnaismuuttujaa X, jos

$$\lim_{i\to\infty} F_i(x) = G(x)$$

pätee kaikissa pisteissä $x \in \mathbb{R}$, joissa kertymäfunktio G on jatkuva. Jakaumasuppenemista merkitään $X_i \xrightarrow{\mathrm{D}} X$. Jakaumasuppenemisesta saatetaan puhua myös heikkona suppenemisena.

Esitellään seuraavaksi tulos, josta ilmenee millaisessa suhteessa suppenemiset ovat keskenään. Todistus on mahdollista lukea lähteestä Izyurov [9] (Propositio 2.7.5).

Propositio 2.1. *Seuraavat implikaatiot pätevät*

1. *Melkein varmasta suppenemisesta seuraa stokastinen suppeneminen.*
2. *Stokastisesta suppenemisesta seuraa jakaumasuppeneminen.*
3. *Kun $p \geq 1$, L^p suppenemisesta seuraa stokastinen suppeneminen.*

Muotoillaan nyt vahva suurten lukujen laki.

Lause 2.3 (Vahva suurten lukujen laki)**.** *Olkoon X_i riippumattomia, samoin jakautuneita satunnaismuuttujia. Olkoon $\mathbb{E}(X_1^4) < \infty$.*

Tällöin todennäköisyydellä 1, pätee

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X_i).$$

Toisin sanoen

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{m.v.} \mathbb{E}(X_i).$$

2.2.1 Lebesque-Stieltjes integraalista

Olkoon $0 \leq b$ ja olkoon $[0, b] \subset \mathbb{R}$. Olkoon $g : [0, b] \to \mathbb{R}$ kasvava cadlag funktio. Voimme määritellä välille $[0, b]$ mitan m_g, jolle pätee

$$m_g((\alpha, \beta]) = g(\beta) - g(\alpha)$$

kaikilla $\alpha \geq 0$ ja $\beta \in (\alpha, b]$. Oletetaan, että $m_g(\{\alpha\}) = 0$. Olkoon $\mathcal{B}$ joukko, jonka alkoita ovat Borel joukot $B \subset [0, b]$. Oletusten mukaan m_g on mitta, joten reaaliluku $m_g(B)$ toteuttaa seuraavat ehdot:

1. $m_g(\emptyset) = 0$
2. $m_g(B) \geq 0 \qquad$ kaikilla $B \in \mathcal{B}$
3. Jos joukot B_i ovat erillisiä joukkoja, joille pätee $B_i \subset [0, b]$ ja $B_i \in \mathcal{B}$ kaikilla $i = 1, 2, ...$, niin pätee yhtäsuuruus $m_g(\cup_{i=0}^{\infty} B_i) = \sum_{i=0}^{\infty} m_g(B_i)$.

Lähdetään nyt kokoamaan Lebesgue-Stieltjes integraalin määritelmää. Tarkastellaan ensiksi funktiota $f : [0, b] \to \mathbb{R}$, joka on mitallinen ja jolle pätee

$$f(t) = \sum_{i=1}^{n} a_i \mathbb{1}(t \in B_i), \qquad \text{missä } a_i \geq 0, B_i \in \mathcal{B} \text{ ja joukot } B_i \text{ osittavat välin } [0, b]. \tag{2.6}$$

Tällöin määrittelemme Lebesgue-Stieltjes integraalin

$$\int_{(0,b]} f(t)\, \mathrm{d}g(t) = \sum_{i=1}^{n} a_i m_g(B_i).$$

Määritellään vielä Lebesgue-Stieltjes integraali yleisemmälle f. Olkoon $f \geq 0$ ja oletetaan, että on olemassa jono funktioita f_n, joille pätee ominaisuus (2.6). Oletetaan, että $f_n \uparrow f$ pisteittäin, kun $n \to \infty$. Määritellään nyt Lebesgue-Stieltjes integraali

$$\int_{(0,b]} f(t)\, \mathrm{d}g(t) = \lim_{n \to \infty} \int_{(0,b]} f_n(t)\, \mathrm{d}g(t).$$

Ehto $f \geq 0$ ei aina päde, mutta integraali halutaan silti määritellä. Tällöin funktio f voidaan jakaa positiiviseen $f^+(t) = f(t)\mathbb{1}(f(t) \geq 0)$ osaan ja vastaavasti negatiiviseen $f^-(t) = -f(t)\mathbb{1}(f(t) \leq 0)$ osaan sekä määritellä integraali seuraavasti

$$\int_{(0,b]} f(t)\, \mathrm{d}g(t) = \int_{(0,b]} f^+(t)\, \mathrm{d}g(t) - \int_{(0,b]} f^-(t)\, \mathrm{d}g(t).$$

Jotta integraali on hyvin määritelty, täytyy oikean puolen molempien integraalien oltava määriteltyjä.

Lisätietoja Lebesgue-Stieltjes integraalista löytyy lähteestä Lehtomaa [10].

2.2.2 Markov prosessista

Määritellään matemaattinen ympäristö, jossa vakuutusyhtiö voi mallintaa kuolevuus ongelmaa. Kuten kilpailevien kuolinsyiden kappaleessa kuvassa 2.1, käytetään ongelman kuvaamiseen vakuutetun mahdollisia eri tiloja. Tila voi olla vaikka 'elossa' tai 'kuollut kuolinsyyhyn h'. Äärellis-tilaiset Markov prosessit osoittautuvat järkeväksi tavaksi mallintaa tätä tilannetta.

Määritelmä 2.16 (Markov prosessi)**.** Olkoon X stokastinen prosessi ja olkoon $u > t \geq 0$. Jos pätee

$$\mathbb{P}\Big(X(u) = k \ \mid \ \sigma\Big(X(s), s \leq t\Big)\Big) = \mathbb{P}\Big(X(u) = k \ \mid \ \sigma\Big(X(t)\Big)\Big),$$

niin prosessia X kutsutaan Markov prosessiksi.

Merkitään prosessin tilojen joukkoa $S = \{0, 1, ..., k\}$. Oletetaan, että $\mathbb{P}(X(t) \in S) = 1$ kaikilla $t \geq 0$. Koska Markov-prosessin tilojen välillä voi olla mahdollista siirtyä, on järkevää määrittää tilasta j tilaan k siirtymisen todennäköisyys. Kutsutaan näitä todennäköisyyksiä siirtymätodennäköisyyksiksi.

Määritelmä 2.17 (Siirtymätodennäköisyys)**.** Olkoon $0 \leq t \leq u$. Oletetaan lisäksi, että $j, k \in S$. Todennäköisyyksiä

$$P_{jk}(t, u) = \mathbb{P}\Big(X(u) = k \ \mid \ X(t) = j\Big)$$

kutsutaan siirtymätodennäköisyyksiksi.

2.2.3 Martingaalista

Määritellään aluksi ehdollinen odotusarvo ja esitellään sitten muutama ehdollisen odotusarvon ominaisuus.

Määritelmä 2.18 (Ehdollinen odotusarvo)**.** Olkoon $(\Omega, \mathscr{F}, \mathbb{P})$ todennäköisyysavaruus ja olkoon X satunnaismuuttuja, jolle pätee $\mathbb{E}(|X|) < \infty$. Oletetaan, että $\mathscr{G} \subset \mathscr{F}$ on sigma-algebra. Tällöin määritellään, että $\mathbb{E}(X|\mathscr{G})$ on mikä tahansa satunnaismuuttuja, joka on $\mathscr{G}$-mitallinen ja jolle pätee millä tahansa $A \in \mathscr{G}$

$$\mathbb{E}\Big(\mathbb{1}_A \mathbb{E}(X|\mathscr{G})\Big) = \mathbb{E}(\mathbb{1}_A X).$$

Esitellään seuraavaksi muutama tärkeä ehdollisen odotusarvon ominaisuus, joita tullaan tässä työssä käyttämään.

Propositio 2.2. *Ehdolliselle odotusarvolle pätee seuraavat ominaisuudet:*

1. *Jos satunnaismuuttuja X on mitallinen sigma-algebran $\mathscr{G}$ suhteen, pätee*

$$\mathbb{E}(X|\mathscr{G}) = X \qquad \textit{melkein varmasti.}$$

2. *Jos satunnaismuuttuja X on riippumaton sigma-algebrasta $\mathscr{G}$, pätee*

$$\mathbb{E}(X|\mathscr{G}) = \mathbb{E}(X) \qquad \textit{melkein varmasti.}$$

3. *(Iteroitu odotusarvo) Jos pätee $\mathscr{G}_1 \subset \mathscr{G}_2$, niin tällöin pätee*

$$\mathbb{E}\Big(\mathbb{E}(X|\mathscr{G}_2)|\mathscr{G}_1\Big) = \mathbb{E}(X|\mathscr{G}_1).$$

Proposition 2.2. todistus.

Kohtien 1. ja 2. todistus: Oletetaan, että $\mathscr{G} = \mathscr{F}$, toisin sanoen meillä on kaikki tieto tapahtuman lopputuloksesta. Selvästi koska $\mathscr{G} = \mathscr{F}$, niin X on $\mathscr{G}$-mitallinen ja

$$\mathbb{E}(\mathbb{1}_A X) = \mathbb{E}(\mathbb{1}_A X).$$

Sama päättely pätee milloin tahansa X on $\mathscr{G}$-mitallinen.

Oletetaan sitten, että $\mathscr{G} = \{\emptyset, \Omega\}$, toisin sanoen meillä ei ole mitään tietoa tapahtuman lopputuloksesta. Tässä tapauksessa ainoastaan vakiofunktiot ovat $\mathscr{G}$-mitallisia ja koska mikä tahansa vakiofunktio on mitallinen mitä tahansa sigma-algebraa vastaan, vakio $\mathbb{E}(X)$ on $\mathscr{G}$-mitallinen. Lisäksi

$$\mathbb{E}\Big(\mathbb{1}_\emptyset \mathbb{E}(X)\Big) = 0 = \mathbb{E}(\mathbb{1}_\emptyset X) \qquad \text{ja} \qquad \mathbb{E}\Big(\mathbb{1}_\Omega \mathbb{E}(X)\Big) = \mathbb{E}\Big(\mathbb{E}(X)\Big) = \mathbb{E}(X) = \mathbb{E}(\mathbb{1}_\Omega X).$$

Oletetaan lopuksi, että X on riippumaton sigma-algebrasta $\mathscr{G}$. Tällöin X ja $\mathbb{1}_A$ ovat riippumattomia kaikilla $A \in \mathscr{G}$. Koska $\mathbb{E}(X)$ on vakio, on se mitallinen kaikkien sigma-algebroiden $\mathscr{G}$ suhteen. Lisäksi riippumattomuuden nojalla pätee

$$\mathbb{E}\Big(\mathbb{1}_A \mathbb{E}(X)\Big) = \mathbb{E}(\mathbb{1}_A)\mathbb{E}(X) = \mathbb{E}(\mathbb{1}_A X).$$

Kohdan 3. todistus: Löytyy lähteestä Izyurov [9] (Propositio 4.2.1). □

Muodostetaan seuraavaksi martingaalin ja alimartingaalin käsitteet, jotka osoittautuvat tämän työn kannalta erittäin keskeisiksi käsitteiksi.

Määritelmä 2.19 (Martingaali)**.** Olkoon M cadlag prosessi. Prosessia M kutsutaan martingaaliksi sigma-algebran $\mathscr{F}_t$ suhteen, mikäli seuraavat ehdot täyttyvät

1. M on sopiva prosessi sigma-algebran $\mathscr{F}_t$ suhteen
2. $\mathbb{E}(|M(t)|) < \infty \qquad$ kaikilla $t \in [0, b]$
3. $\mathbb{E}\Big(M(t)|\mathscr{F}_s\Big) = M(s) \qquad$ kaikilla $s \leq t$

Ehtoa 3 kutsutaan martingaali-ominaisuudeksi. Jos ehto 3 korvataan epäyhtälöllä

$$\mathbb{E}\Big(M(t)|\mathscr{F}_s\Big) \geq M(s) \qquad \text{kaikilla } s \leq t,$$

prosessia M kutsutaan alimartingaaliksi. Osoitetaan seuraavaksi, että martingaalin odotusarvo pysyy kaikilla ajanhetkillä samana.

Lemma 2.1. *Olkoon M martingaali, kuten edellä.*
Tällöin parametrilla $t > 0$ pätee

$$\mathbb{E}\Big(M(t)\Big) = \mathbb{E}\Big(M(0)\Big).$$

Lemman 2.1. todistus. Käyttämällä martingaali-ominaisuutta

$$\mathbb{E}\Big(M(t)\Big) = \mathbb{E}\Big(\mathbb{E}\Big(M(t)|\mathscr{F}_0\Big)\Big) = \mathbb{E}\Big(M(0)\Big).$$

□

Huomautus 2.5. Erityisesti, kun $M(0) = 0$ pätee $\mathbb{E}(M(t)) = 0$ kaikilla t.

Määritellään vielä neliöllisesti integroituva martingaali. Melkein kaikki tässä työssä käsitellyt martingaalit ovat neliöllisesti integroituvia, joten käsite on varsin keskeinen.

Määritelmä 2.20 (Neliöllisesti integroituva martingaali)**.** Martingaalin M sanotaan olevan neliöllisesti integroituva, jos

$$\sup_{t\in[0,b]} \mathbb{E}\Big(M(t)^2\Big) < \infty.$$

Käsittelemme tässä työssä myös paljon lokaaleja martingaaleja, joten mainitaan niistä muutama seikka.

Huomautus 2.6. Martingaali on aika lokaali martingaali. Päinvastainen ei kuitenkaan päde.

Huomautus 2.7. Kahden lokaalin martingaalin summa on lokaali martingaali. Samoin kahden lokaalin martingaalin erotus on lokaali martingaali.

2.2.4 Neliöllisestä varianssista ja kovarianssista

Määritellään tässä kappaleessa neliöllinen varianssi kuten Lehtomaa[10]. Määritellään sitä varten ensin rajoitetusti heilahteleva kuvaus.

Määritelmä 2.21 (Rajoitetusti heilahteleva kuvaus)**.** Olkoon $[0,b] \subset \mathbb{R}$. Funktion $f : [0,b] \to \mathbb{R}$ sanotaan olevan rajoitetusti heilahteleva, mikäli

$$\sup\left\{ \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)| \;\middle|\; 0 = t_0 < t_1 < ... < t_{n-1} < t_n = b, n \in \mathbb{N} \right\} < \infty$$

Nyt voimme määritellä neliöllisen varianssin seuraavasti:

Määritelmä 2.22 (Neliöllinen varianssi)**.** Oletetaan, että f on rajoitetusti heilahteleva, cadlag funktio. Funktion f neliöllinen varianssi välillä $[0,b]$ määritellään olevan

$$[f] = \lim_{n\to\infty} \sum_{i=0}^{n-1} \Big(f(t_{i+1}) - f(t_i) \Big)^2,$$

missä $0 = t_0 < t_1 < ... < t_{n-1} < t_n = b$ ja $\max\Big\{ t_{i+1} - t_i \;\mid\; i \in \{0, 1, ..., n-1\} \Big\} \to 0$, kun $n \to \infty$.

Huomautus 2.8. Vain funktion epäjatkuvat kohdat vaikuttavat neliöllisen varianssin summaan, sillä jatkuvuuden määritelmän nojalla summaan lisätään nollan suuruisia termejä jatkuvissa kohdissa.

Huomautus 2.9. Kun puhutaan prosessin neliöllisestä varianssista, tarkoitetaan prosessin polkujen neliöllistä varianssia, joka on prosessi jos polku itsessään on satunnainen.

Esitellään seuraavaksi toinen tapa neliöllisen varianssin laskemiseen. Lemman todistus löytyy lähteestä Lehtomaa [10] (Lemma 1.6).

Lemma 2.2. *Oletetaan, että funktio f on cadlag funktio, joka on rajoitetusti heilahteleva välillä $[0,b]$.*

Tällöin

$$[f] = f(b)^2 - 2\int_0^b f(t-)\,\mathrm{d}f(t).$$

Voimme vastaavasti määritellä kahden funktion kovarianssin.

Määritelmä 2.23 (Kovarianssi)**.** Oletetaan, että cadlag funktiot f ja g ovat välillä $[0, b]$ rajoitetusti heilahtelevia. Funktioiden f ja g kovarianssi välillä $[0, b]$ määritellään olevan

$$[f, g] = \lim_{n\to\infty} \sum_{i=0}^{n-1} \Big(f(t_{i+1}) - f(t_i)\Big)\Big(g(t_{i+1}) - g(t_i)\Big),$$

missä $0 = t_0 < t_1 < ... < t_{n-1} < t_n = b$ ja $\max\Big\{t_{i+1} - t_i \mid i \in \{0, 1, ..., n-1\}\Big\} \to 0$, kun $n \to \infty$.

Määritellään vielä lopuksi äärellisesti heilahteleva prosessi.

Määritelmä 2.24 (Äärellisesti heilahteleva prosessi)**.** Prosessin X sanotaan olevan äärellisesti heilahteleva, jos se on rajoitetusti heilahteleva kaikilla äärellisillä aikavälillä todennäköisyydellä 1.

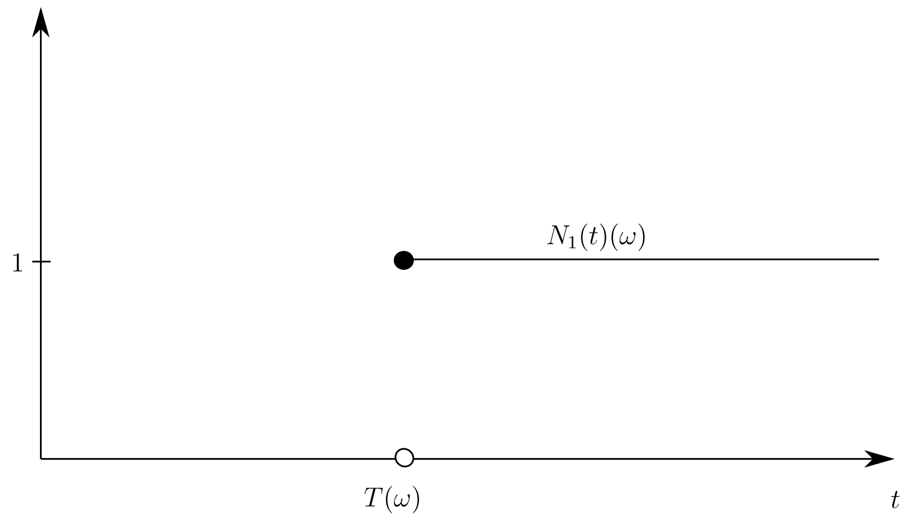
2.2.5 Laskuriprosesseista

Määritellään aluksi laskuriprosessi ja havainnollistetaan sitten määritelmää esimerkin ja kuvan avulla.

Määritelmä 2.25 (Moniulotteinen laskuriprosessi)**.** Moniulotteinen laskuriprosessi $N = (N_1, ..., N_k)$ on vektori, jonka komponentit ovat sopivia cadlag prosesseja, joiden polut ovat paloittain vakiota ja kasvavia. Polkujen hyppyjen ainoa sallittu koko on $+1$. Lisäksi pätee $N_h(0) = 0$, missä $h = 1, ..., k$. Vaadimme lisäksi, että mitkään kaksi komponenttia eivät hyppää samanaikaisesti.

Huomautus 2.10. Moniulotteisen laskuriprosessin komponentit ovat itsessään laskuriprosesseja.

Intuitiivisesti laskuriprosessin voidaan ajatella kuvata lukumäärän laskemista. Yksinkertainen esimerkki laskuriprosessita voisi olla Markov prosessi, jossa on kaksi tilaa: 0 ja 1. Tilalla 0 tarkoitetaan, että vakuutettu on elossa ja tilalla 1, että vakuutettu on kuollut. Merkitään $N_1(t) = \mathbb{1}(T \leq t)$. Tällöin laskuriprosessi N_1 saa arvon 0, kun vakuutettu on elossa ja arvon 1 kuolemasta eteenpäin. Laskuriprosessin voi siis ajatella laskevan siirtymiä Markov prosessin tilasta toiseen.



Kuva 2.2: Laskuriprosessin N_1 polku.

Esitetään seuraavaksi esimerkki kilpailevien kuolinsyiden teorian mukaisesta asetelmasta, jossa on n vakuutettua ja k eri kuolinsyytä. Osoitetaan, että tässä tilanteessa tarkastellut laskuriprosessit ovat alimartingaaleja.

Esimerkki 2.3. Oletetaan kilpailevien kuolinsyiden alainen teoria, missä vakuutettuja on n yksilöä ja mahdollisia eri kuolinsyitä k kappaletta. Oletetaan, että kaikki vakuutetut ovat x-ikäisiä ajanhetkellä 0. Mallinnetaan vakuutetun i tilaa Markov prosessilla Z_i, missä

$$Z_i(t) = \sum_{h=1}^{k} h\mathbb{1}\Big(T^i(x) \leq t, \text{ kuolinsyy on } h\Big).$$

Oletetaan, että filtraatio on muotoa $\mathscr{F}_t = \sigma(Z_i(s) \,|\, s \leq t, i = 1, ..., n)$. Oletetaan lisäksi, että on olemassa k-ulotteinen laskuriprosessi N, jonka komponentit ovat muotoa

$$N_h(t) = \sum_{i=1}^{n} \mathbb{1}\Big(T^i(x) \leq t, \text{kuolinsyy on } h\Big).$$

Nyt N_h on sopiva prosessi filtraation $\mathscr{F}_t$ suhteen. Lisäksi $\mathbb{E}(|N_h(t)|) < \infty$, sillä kaikilla $t \in [0, b]$ pätee $N_h(t) \leq n$. Lopuksi koska N_h on kasvava, pätee $N_h(t) \geq N_h(s)$ kaikilla $s \leq t$, jolloin

$$\mathbb{E}\Big(N_h(t) \,|\, \mathscr{F}_s\Big) \geq \mathbb{E}\Big(N_h(s) \,|\, \mathscr{F}_s\Big) = N_h(s).$$

Täten N_h on alimartingaali.

2.2.6 Kompensaattoreista

Beiglboeckin, Schachermayerin ja Veliyevin artikkeli *A short proof of the Doob-Meyer theorem* [4] osoittaa, että alimartingaali voidaan kirjoittaa martingaalin M sekä kasvavan, ennustettavan prosessin Λ summana. Täten saamme tuloksen, joka kertoo hyödyllisen esitystavan esimerkin 2.3 laskuriprosessille, jossa kyseinen prosessi on jaettu systemaattiseen (kompensaattori) ja täysin satunnaiseen (martingaali) osaan. Esitellään aluksi hajotelma sen yleisessä esitystavassa ja sen jälkeen lokalisointia hyödyntävässä muodossa, jota voidaan soveltaa laajemmalle joukolle prosesseja.

Lause 2.4 (Doob-Meyer hajotelma)**.** *Oletetaan, että $S = (S_t)_{0\leq t\leq b}$ on cadlag alimartingaali, jolle pätee, että joukko satunnaismuuttujia S_τ on tasaisesti integroituva. Parametri τ on mielivaltainen pysäytyshetki. Tällöin on olemassa yksikäsitteinen tapa kirjoittaa S muodossa*

$$S = M + \Lambda,$$

missä M on martingaali ja Λ on ennustettava ja kasvava prosessi, jolle pätee $\Lambda(0) = 0$.

Intuitiivisen käsityksen kompensaattorista saa muodosta

$$\mathrm{d}\Lambda(t) = \mathbb{E}\Big(\mathrm{d}S(t)|\mathscr{F}_{t-}\Big).$$

Kutsumme prossessia λ_h laskuriprosessin N_h intensiteettiprosessiksi, jos λ_h on ennustettava prosessi ja pätee

$$\Lambda_h(t) = \int_0^t \lambda_h(s)\,\mathrm{d}s \qquad \text{kaikilla } t.$$

Lähteen Andersen et al. [3] luvun II mukaan on olemassa tapa kirjoittaa Doob-Meyer hajotelma ilman integoituvuus-ehtoa käyttämällä lokalisointia. Oletetaan, että S on cadlag, sopiva prosessi. Sanotaan, että Λ on prosessin S kompensaattori, jos Λ on ennustettava, cadlag ja äärellisesti heilahteleva prosessi siten, että $M = S - \Lambda$ on lokaali martingaali ja $M(0) = 0$. Mikäli kompensaattori on olemassa on se yksikäsitteinen.

Lisäksi lähteen Andersen et al. [3] mukaan prosessilla S on kompensaattori jos ja vain jos S on kahden lokaalin alimartingaalin erotus.

Olkoon M ja M' lokaaleja neliöllisesti integroituvia martingaaleja. Tällöin Jensenin epäyhtälön nojalla M^2 on lokaali alimartingaali, sillä

$$\mathbb{E}\Big(M^2(t)|\mathscr{F}_s\Big) \geq \Big(\mathbb{E}\Big(M(t)|\mathscr{F}_s\Big)\Big)^2 = M^2(s).$$

Doob-Meier hajotelman mukaan prosessilla M^2 on tällöin kompensaattori. Merkitään sitä $\langle M, M\rangle$ tai lyhyesti $\langle M\rangle$. Englanninkielisessä kirjallisuudessa termistä $\langle M\rangle$ puhutaan nimellä *predictable variation process*.

Edellisistä huomioista ja lokaalista Doob-Meyer hajotelmasta koottuna saadaan seuraava huomautus.

Huomautus 2.11. Olkoon M lokaali neliöllisesti integroituva martingaali. Tällöin $M^2 - \langle M \rangle$ on lokaali martingaali, joka saa arvon 0 ajanhetkellä 0.

Toisaalta koska voidaan kirjoittaa

$$MM' = \frac{1}{4}(M + M')^2 - \frac{1}{4}(M - M')^2,$$

missä edellisen nojalla MM' on kahden lokaalin alimartingaalin erotus, jolloin sillä on kompensaattori. Kompensaattoria merkitään $\langle M, M' \rangle$ ja siitä puhutaan kirjallisuudessa nimellä *predictable covariation process*.

Intuitiivisesti voidaan nyt ajatella

$$\mathrm{d}\langle M \rangle(t) = \mathrm{Var}\Big(\mathrm{d}M(t) | \mathscr{F}_{t-}\Big)$$

ja vastaavasti

$$\mathrm{d}\langle M, M' \rangle(t) = \mathrm{Cov}\Big(\mathrm{d}M(t), \mathrm{d}M'(t) | \mathscr{F}_{t-}\Big).$$

Huomautus 2.12. Jos $M = (M_1, ..., M_k)$ on k-ulotteinen vektori, jonka komponentit ovat lokaaleja neliöllisesti integroituvia martingaaleja, niin $\langle M \rangle$ on $k \times k$ matriisi, jonka komponentit ovat $\langle M_h, M_j \rangle$, missä $h, j = 1, ..., k$.

Tässä työssä tullaan tarkastelemaan paljon Lebesque-Stieltjes integraaleja. Erityistä huomiota kiinnitämme muotoa $\int H \,\mathrm{d}M$ oleviin integraaleihin, joissa H on ennustettava prosessi ja M lokaali neliöllisesti integroituva martingaali. Esitetään seuraavaksi kyseistä integraalia koskeva tulos.

Lemma 2.3. *Olkoon M lokaali neliöllisesti integroituva, äärellisesti heilahteleva martingaali. Olkoon H ennustettava prosessi. Oletetaan lisäksi, että joko $\int H^2 \,\mathrm{d}[M]$ on lokaalisti integroituva tai $\int H^2 \,\mathrm{d}\langle M \rangle$ on lokaalisti äärellinen.*

Tällöin $\int H \,\mathrm{d}M$ on lokaali neliöllisesti integroituva martingaali ja

$$\left[\int H \,\mathrm{d}M\right] = \int H^2 \,\mathrm{d}[M] \qquad \text{ja}$$

$$\left\langle \int H \,\mathrm{d}M \right\rangle = \int H^2 \,\mathrm{d}\langle M \rangle.$$

Huomautus 2.13. Ehto: $\int H^2 \,\mathrm{d}\langle M \rangle$ on lokaalisti äärellinen, on totta, jos H on lokaalisti rajoitettu. Ennustettava prosessi H on puolestaan lokaalisti rajoitettu, jos sen polut ovat vasemmalta jatkuvia ja oikeat raja-arvot on olemassa.

Tutustutaan seuraavassa esimerkissä erään laskuriprosessin kompensaattoriin. Esimerkin 2.3 laskuriprosessi N_h on alimartingaali, joten sen voi esittää kompensaattorin ja martingaalin summana. Käytämme tätä tulosta myös myöhemmin tässä työssä hyvin keskeisessä roolissa.

Esimerkki 2.4. Oletetaan, että vakuutuskannassa on yksi vakuutettu ja oletetaan, että vakuutetun ikä ajanhetkellä 0 on x. Olkoon $T(x)$ positiivinen satunnaismuuttuja, joka kuvaa vakuutetun jäljellä olevaa elinaikaa. Koska kyseessä on elinaika, satunnaismuuttujan $T(x)$ on saatava vain äärellisiä arvoja, toisin sanoen $\mathbb{P}(T(x) < \infty) = 1$. Mallinnetaan jälleen vakuutetun tilaa Markov prosessilla Z, missä

$$Z(t) = \sum_{h=1}^{k} h \mathbb{1}\Big(T(x) \leq t, \text{ kuolinsyy on } h\Big).$$

Määritellään k-ulotteinen laskuriprosessi

$$N(t) = \Big(N_1(t), ..., N_k(t)\Big),$$

missä $N_h(t) = \mathbb{1}(T(x) \leq t, \text{ kuolinsyy on } h)$, kun $h \in \{1, ..., k\}$. Olkoon filtraatio muotoa

$$\mathscr{F}_t = \sigma\Big(Z(s) \mid s \leq t\Big).$$

Olkoon τ_F yläraja funktion F_x kantajasta. Oletetaan, että

$$\int_0^t \mu_h^x(s) \, \mathrm{d}s < \infty$$

kaikilla $t < \tau_F$. Tällöin laskuriprosessilla N on kompensaattori Λ, joka on muotoa

$$\Lambda_h(t) = \int_0^t \mathbb{1}\Big(T(x) \geq s\Big) \mu_h^x(s) \, \mathrm{d}s$$

ja laskuriprosessin N_h intensitettiprosessi λ_h on muotoa

$$\lambda_h(t) = \mu_h^x(t) \mathbb{1}\Big(T(x) \geq t\Big).$$

Todistetaan tämä tulos. Esimerkin 2.3 perusteella laskuriprosessi N_h on alimartingaali, joten se voidaan hajoittaa Doob-Meier hajotelman nojalla kompensaattorin ja martingaalin summaksi. Laskuriprosessin määritelmän mukaan N_h on cadlag, sopiva prosessi. Koska Λ_h on jatkuva ja sopiva prosessi, on se myös ennustettava. Jatkuvuudesta seuraa myös cadlag ominaisuus. Lisäksi Λ_h on äärellisesti heilahteleva prosessi. Nyt myös pätee, että $M_h(0) = N_h(0) - \Lambda_h(0) = 0$. Täten esimerkin todistamiseksi riittää osoittaa, että

$$M_h(t) = N_h(t) - \Lambda_h(t) = \mathbb{1}\Big(T(x) \leq t, \text{ kuolinsyy on } h\Big) - \int_0^t \mathbb{1}\Big(T(x) \geq s\Big) \mu_h^x(s) \, \mathrm{d}s$$

on martingaali, jolloin se on myös lokaali martingaali. Nyt $M_h(t)$ on $\mathscr{F}_t$-mitallinen ja $\mathbb{E}(|M_h(t)|) < \infty$, joten osoitettavaksi jää ainoastaan martingaaliominaisuus.

Olkoon $s < t$. Parametrin t kasvaessa rajatta jokaisella $\omega \in \Omega$ martingaalin $M_h(t)$ arvo suppenee kohti jotain vakiota. Merkitään tätä arvoa $M_h(\infty)$. Jos voimme osoittaa, että

$$\mathbb{E}\Big(M_h(\infty)|\mathscr{F}_t\Big) = M_h(t),$$

saamme todistettua martingaali-ominaisuuden käyttämällä iteroitua odotusarvoa

$$\begin{aligned}
\mathbb{E}\Big(M_h(t)|\mathscr{F}_s\Big) &= \mathbb{E}\Big(\mathbb{E}\Big(M_h(\infty)|\mathscr{F}_t\Big)|\mathscr{F}_s\Big) \\
&= \mathbb{E}\Big(M_h(\infty)|\mathscr{F}_s\Big) \\
&= M_h(s).
\end{aligned}$$

Oletetaan aluksi, että $t = 0$. Nyt $\mathscr{F}_0 = \{\emptyset, \Omega\}$ ja $M_h(0) = 0$, joten meidän täytyy todistaa ainoastaan

$$\mathbb{E}\Big(M_h(\infty)\Big) = 0.$$

Nyt $\mathbb{1}(T(x) \geq s)\mu_h^x(s)$ on positiivinen funktio, joten käyttämällä Fubinin lausetta ja lausetta 2.2 saadaan

$$\begin{aligned}
\mathbb{E}\Big(M_h(\infty)\Big) &= \mathbb{E}\Big(N_h(\infty)\Big) - \mathbb{E}\Big(\Lambda_h(\infty)\Big) \\
&= \mathbb{E}\Big(\mathbb{1}\Big(T(x) \leq \infty,\ \text{kuolinsyy on } h\Big)\Big) - \mathbb{E}\left(\int_0^{\tau_F} \mathbb{1}\Big(T(x) \geq s\Big)\mu_h^x(s)\,\mathrm{d}s\right) \\
&= \mathbb{P}\Big(T(x) \leq \infty,\ \text{kuolinsyy on } h\Big) - \int_0^{\tau_F} \mathbb{P}\Big(T(x) \geq s\Big)\mu_h^x(s)\,\mathrm{d}s \\
&= G_h(\infty) - \int_0^{\tau_F} G_h^{'}(s)\,\mathrm{d}s \\
&= G_h(\infty) - G_h(\tau_F) + G_h(0) \\
&= 0,
\end{aligned}$$

sillä $G_h(0) = 0$ ja $G_h(\infty) = G_h(\tau_F)$. Täten saatiin osoitettua, että martingaali-ominaisuus pätee, kun $t = 0$.

Osoitetaan tulos seuraavaksi mielivaltaiselle t. Jaetaan todistus kahteen osaan: oletetaan ensin, että $T(x) = s \leq t$ jollekin $s \leq t$ ja seuraavaksi $T(x) > t$.

Tapaus 1: Oletetaan, että $T(x) = s \leq t$ jollekin $s \leq t$. Nyt

$$\begin{aligned}
M_h(\infty) &= N_h(\infty) - \Lambda_h(\infty) \\
&= \mathbb{1}\Big(T(x) \leq \infty,\ \text{kuolinsyy on } h\Big) - \int_0^{\tau_F} \mathbb{1}\Big(T(x) \geq u\Big)\mu_h^x(u)\,\mathrm{d}u \\
&= \mathbb{1}(\text{kuolinsyy on } h) - \int_0^{s} \mu_h^x(u)\,\mathrm{d}u
\end{aligned}$$

ja

$$\begin{aligned}M_h(t) &= N_h(t) - \Lambda_h(t)\\ &= \mathbb{1}\Big(T(x) \leq t,\ \text{kuolinsyy on } h\Big) - \int_0^t \mathbb{1}\Big(T(x) \geq u\Big)\mu_h^x(u)\,\mathrm{d}u\\ &= \mathbb{1}(\text{kuolinsyy on } h) - \int_0^s \mu_h^x(u)\,\mathrm{d}u\end{aligned}$$

ja

$$\begin{aligned}M_h(s) &= N_h(s) - \Lambda_h(s)\\ &= \mathbb{1}\Big(T(x) \leq s,\ \text{kuolinsyy on } h\Big) - \int_0^s \mathbb{1}\Big(T(x) \geq u\Big)\mu_h^x(u)\,\mathrm{d}u\\ &= \mathbb{1}(\text{kuolinsyy on } h) - \int_0^s \mu_h^x(u)\,\mathrm{d}u.\end{aligned}$$

Saadaan, että $M_h(\infty) = M_h(t) = M_h(s)$. Tällöin

$$\mathbb{E}\Big(M_h(\infty)|\mathscr{F}_t\Big) = \mathbb{E}\Big(M_h(t)|\mathscr{F}_t\Big) = M_h(t),$$

sillä $M_h(t)$ on $\mathscr{F}_t$-mitallinen. Samoin

$$\mathbb{E}\Big(M_h(t)|\mathscr{F}_s\Big) = \mathbb{E}\Big(M_h(s)|\mathscr{F}_s\Big) = M_h(s).$$

Tapaus 2: Oletetaan seuraavaksi, että $T(x) > t$. Koska yhtälö

$$\mathbb{E}\Big(M_h(\infty)|\mathscr{F}_t\Big) = M_h(t)$$

voidaan kirjoittaa muodossa

$$\begin{aligned}\mathbb{E}\Big(M_h(\infty) - M_h(t)|\mathscr{F}_t\Big) &= \mathbb{E}\Big(M_h(\infty)|\mathscr{F}_t\Big) - \mathbb{E}\Big(M_h(t)|\mathscr{F}_t\Big)\\ &= \mathbb{E}\Big(M_h(\infty)|\mathscr{F}_t\Big) - M_h(t)\\ &= 0,\end{aligned}$$

meidän täytyy osoittaa $\mathbb{E}(M_h(\infty) - M_h(t)|\mathscr{F}_t) = 0$. Nyt

$$\begin{aligned}\mathbb{E}\Big(M_h(\infty) - M_h(t)|\mathscr{F}_t\Big) &= \mathbb{E}\Big(N_h(\infty) - \Lambda_h(\infty) - N_h(t) + \Lambda_h(t)|\mathscr{F}_t\Big)\\ &= \mathbb{E}\Big(N_h(\infty) - N_h(t)|\mathscr{F}_t\Big) - \mathbb{E}\Big(\Lambda_h(\infty) - \Lambda_h(t)|\mathscr{F}_t\Big).\end{aligned}$$

Kun $T(x) > t$, pätee

$$\begin{aligned}
N_h(\infty) - N_h(t) &= \mathbb{1}\Big(T(x) \leq \infty,\ \text{kuolinsyy on } h\Big) - \mathbb{1}\Big(T(x) \leq t,\ \text{kuolinsyy on } h\Big) \\
&= \mathbb{1}\Big(T(x) \leq \infty,\ \text{kuolinsyy on } h\Big) - 0 \\
&= \mathbb{1}\Big(T(x) \leq \infty,\ \text{kuolinsyy on } h\Big).
\end{aligned}$$

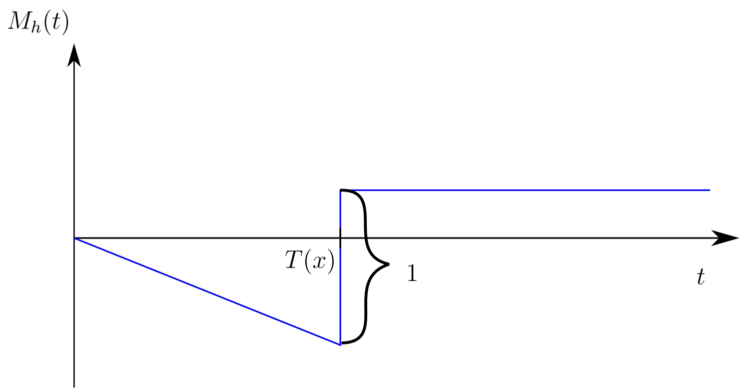
Lisäksi

$$\begin{aligned}
\Lambda_h(\infty) - \Lambda_h(t) &= \int_0^{\tau_F} \mathbb{1}\Big(T(x) \geq s\Big)\mu_h^x(s)\,\mathrm{d}s - \int_0^t \mathbb{1}\Big(T(x) \geq s\Big)\mu_h^x(s)\,\mathrm{d}s \\
&= \int_t^{\tau_F} \mathbb{1}\Big(T(x) \geq s\Big)\mu_h^x(s)\,\mathrm{d}s.
\end{aligned}$$

Täten saadaan

$$\begin{aligned}
\mathbb{E}\Big(M_h(\infty) - M_h(t)|\mathscr{F}_t\Big) = \mathbb{P}\Big(T(x) \leq \infty,\ \text{kuolinsyy on } h\Big) \\
- \mathbb{E}\Big(\int_0^{\tau_F} \mathbb{1}\Big(T(x) \geq u\Big)\mu_h^x(u)\mathbb{1}_{(t,\tau_F)}\,\mathrm{d}u\Big),
\end{aligned}$$

kun $T(x) > t$. Kuolevuus on nyt muotoa $\mu_h^x \mathbb{1}_{(t,\tau_F)}$. Täten olemme samassa tilanteessa kuin $t = 0$, mutta kuolevuus on eri. Todistus etenee kuin kohdassa $t = 0$.



Kuva 2.3: Sinisellä piirretty esimerkki martingaalin M_h polusta tapauksessa, jossa kuolevuus on vakio. Martingaali on vähenevä kuolinhetkeen asti, jolloin se hyppää ylöspäin arvon 1 verran ja pysyy tässä vakioarvossa kuolemasta eteenpäin.

3 Nelson-Aalen estimaattorista

3.1 Estimaattorin muodostaminen

Seuraamme tässä luvussa lähteen Andersen et al. [3] luvun IV. Nonparametric estimation merkintä- ja esitystapaa.

Tarkastellaan tilannetta, jossa on n samankaltaista vakuutettua sekä vakuutetun i mahdollisia kuolinsyitä k kappaletta, kun $i = 1, ..., n$. Oletetaan parametrin n arvon olevan suhteellisen suuri ja parametrin k arvon suhteellisen pieni. Oletetaan, että kaikki vakuutetut ovat samanikäisiä, toisin sanoen oletetaan, että ajanhetkellä 0 kaikki vakuutetut ovat x-ikäisiä. Oletetaan vakuutettujen jäljellä olevat elinajat $T^i(x)$ eri kuolinsyiden suhteen samoin jakautuneiksi ja oletetaan, että vakuutettujen elinajat ovat toisistaan riippumattomia satunnaismuuttujia. Joudumme lisäksi olettamaan, että vakuutetun kuolinsyy on yksiselitteisesti määriteltävissä, eikä kuolema johdu monesta tekijästä. Tehdään vielä oletus, että vain yksi vakuutettu voi kuolla kerrallaan. Kuvataan mainittua tilannetta monitilaisena, äärellisenä Markov prosessina, jossa tila 0 vastaa tilannetta, että vakuutettu on elossa ja tila h tilannetta, että vakuutettu on kuollut johtuen syystä h, missä $h = 1, ..., k$. Jokainen vakuutettu aloittaa tilasta 0.

Mallinnetaan vakuutetun i tilaa Markov prosessilla Z_i, missä

$$Z_i(t) = \sum_{h=1}^{k} h \mathbb{1}\Big(T^i(x) \leq t, \text{ kuolinsyy on } h\Big).$$

Huomautus 3.1. Jokainen vakuutettu aloittaa tilasta 0 eikä pääse tiloista $1, ..., k$ enää pois, joten käytetään siirtymätodennäköisyyksistä $P_{0k}(0, s)$ lyhennysmerkintää $P_k(0, u)$.

Määritellään laskuriprosessi

$$N_h(t) = \sum_{i=1}^{n} \mathbb{1}\Big(T^i(x) \leq t, \text{kuolinsyy on } h\Big),$$

joka laskee kaikkien vakuutettujen siirtymiä tilasta 0 tilaan h aikavälillä $[0, t]$. Olkoon filtraatio muotoa $\mathscr{F}_t = \sigma(Z_i(s) \,|\, s \leq t, i = 1, ..., n)$. Tavoitteena on estimoida kumulatiivista funktiota

$$A_h(t) = \int_0^t \mu_h^x(s) \, \mathrm{d}s,$$

missä $h = 1, ..., k$. Tiedämme, että jos $n = 1$ laskuriprosessin kompensaattori on esimerkin 2.4 perusteella

$$\Lambda_h(t) = \int_0^t \mu_h^x(s) \mathbb{1}\Big(T(x) \geq s\Big) \, \mathrm{d}s.$$

Kun vakuutettuja on n kappaletta kompensaattori saa odotusarvon lineaarisuuden perusteella muodon

$$\Lambda_h(t) = \sum_{i=1}^{n} \int_0^t \mu_h^x(s) Y_i(s)\,\mathrm{d}s,$$

missä $Y_i(s) = \mathbb{1}(T^i(x) \geq s)$ ja $Y(s) = \sum_{i=1}^n Y_i(s) = \sum_{i=1}^n \mathbb{1}(T^i(x) \geq s)$ on ennustettava prosessi, joka kuvaa riskin kohteena olevien vakuutettujen määrää. Toisin sanoen Y on elossa olevien vakuutettujen määrä. Oletetaan, että

$$\int_0^t \mu_h^x(s)\,\mathrm{d}s < \infty$$

kaikilla h ja kaikilla $t \in [0, b]$. Oletetaan lisäksi, että μ_h^x on rajoitettu välillä $[0, t]$ kaikilla $t \in [0, b]$.

Merkitään

$$M_h(t) = N_h(t) - \Lambda_h(t),$$

missä M_h on lokaali neliöllisesti integroituva martingaali kaikilla $h = 1, ..., k$. Tällöin järjestelemällä termit uudestaan saadaan

$$N_h(t) = M_h(t) + \Lambda_h(t). \tag{3.1}$$

Koska kompensaattorin Λ_h polut ovat jatkuvia ja laskuriprosessin N_h jokaisen hypyn suuruus on 1, pätee $[M_h] = N_h$. Lemman 2.2 nojalla voidaan kirjoittaa

$$[M_h](t) = M_h^2(t) - 2\int_0^t M_h(s-)\,\mathrm{d}M_h(s),$$

jolloin $M_h^2 - [M_h]$ on lokaali martingaali. Lisäksi kompensaattori-kappaleessa todetun nojalla $\langle M_h \rangle$ on prosessin M_h^2 kompensaattori. Saadaan siis

$$[M_h] = M_h^2 + \text{lokaali martingaali}$$

ja

$$\langle M_h \rangle = M_h^2 + \text{lokaali martingaali}.$$

Täten saadaan huomautuksen 2.7 nojalla

$$[M_h] = \langle M_h \rangle + \text{lokaali martingaali},$$

joten $\langle M_h \rangle$ on prosessin $[M_h]$ kompensaattori. Lisäksi koska $[M_h] = N_h$ ja laskuriprosessin N_h kompensaattori on Λ_h, saadaan Doob-Meier hajotelman yksikäsitteisyyden nojalla, että $\langle M_h \rangle = \Lambda_h$.

Lausekkeen (3.1) perusteella voidaan kirjoittaa

$$\begin{aligned}\mathrm{d}N_h(t) &= \mathrm{d}M_h(t) + \mathrm{d}\Lambda_h(t) \\ &= \mathrm{d}M_h(t) + \mu_h^x(t)Y(t)\,\mathrm{d}t.\end{aligned}$$

Yleisesti termiä $\mathrm{d}M_h(t)$ pidetään vain ei-merkittävänä satunnaismeluna, joten funktiolle $A_h(t)$ saadaan estimaattori

$$\hat{A}_h(t) = \int_0^t \frac{1}{Y(s)}\,\mathrm{d}N_h(s).$$

Tätä estimaattoria kutsutaan **Nelson-Aalen estimaattoriksi**. Nollalla jakamisesta ei yleensä aiheudu ongelmia, sillä usein vakuutettujen lukumäärä on niin suuri, että suurella todennäköisyydellä ainakin joku selviää tarkastelukauden loppuun. Tarvittaessa voidaan määrittää $Y(s)^{-1} = 0$, kun $Y(s) = 0$.

Koska laskuriprosessi N_h hyppää arvon 1 verran, kun vakuutettu kuolee kuolinsyyhyn h ja pysyy muuten vakiona, estimaattori $\hat{A}_h(t)$ voidaan kirjoittaa muodossa

$$\sum_{\{i:T^i(x)\leq t\}} \frac{\mathbb{1}(\text{kuolinsyy on } h)}{Y\big(T^i(x)\big)}.$$

Määritellään seuraavaksi indikaattoriprosessi $J(t) = \mathbb{1}(Y(t) > 0)$, missä $J(t)$ saa arvon 0 mikäli kaikki vakuutetut ovat jo kuolleet juuri ennen ajanhetkeä t ja vastaavasti arvon 1 mikäli vakuutettuja on vielä elossa juuri ennen ajanhetkeä t. Määritellään estimaattori

$$A_h^*(t) = \int_0^t \mu_h^x(s)J(s)\,\mathrm{d}s. \tag{3.2}$$

Kun on vain pieni todennäköisyys, että jollekin $s \leq t$ pätee $Y(s) = 0$, estimaattori saa muodon

$$A_h^*(t) \approx \int_0^t \mu_h^x(s)\,\mathrm{d}s = A(t).$$

Koska $J(t) = 1$, vain silloin kun $Y(t)$ on aidosti positiivinen ja laskuriprosessi N_h hyppii vain silloin kun $Y(t)$ on aidosti positiivinen, voimme kirjoittaa Nelson-Aalen estimaattorin $\hat{A}_h(t)$ muotoon

$$\hat{A}_h(t) = \int_0^t \frac{J(s)}{Y(s)}\,\mathrm{d}N_h(s),$$

missä osamäärän on määritelty olevan

$$\frac{J(t)}{Y(t)} = 0,$$

kun $Y(t) = 0$.

Nyt voidaan kirjoittaa lausekkeen (3.1) nojalla

$$\int_0^t \frac{J(s)}{Y(s)}\,\mathrm{d}N_h(s) = \int_0^t \frac{J(s)}{Y(s)}\mu_h^x(s)Y(s)\,\mathrm{d}s + \int_0^t \frac{J(s)}{Y(s)}\,\mathrm{d}M_h(s),$$

josta seuraa

$$\int_0^t \frac{J(s)}{Y(s)}\,\mathrm{d}N_h(s) = \int_0^t J(s)\mu_h^x(s)\,\mathrm{d}s + \int_0^t \frac{J(s)}{Y(s)}\,\mathrm{d}M_h(s)$$

ja edelleen

$$\hat{A}_h(t) = A_h^*(t) + \int_0^t \frac{J(s)}{Y(s)}\,\mathrm{d}M_h(s),$$

joten

$$\hat{A}_h(t) - A_h^*(t) = \int_0^t \frac{J(s)}{Y(s)}\,\mathrm{d}M_h(s). \tag{3.3}$$

3.2 Ominaisuuksia

3.2.1 Odotusarvo

Seurataan lähteen Andersen et al. [3] luvun IV päättelyä ja johdetaan Nelson-Aalen estimaattorin odotusarvo lausekkeesta (3.3).

Oletetaan, että $t \in [0, b]$. Koska Y on ennustettava prosessi, on myös J ennustettava prosessi, jolloin niiden osamäärä J/Y on ennustettava prosessi. Lisäksi J/Y on lokaalisti rajoitettu, sillä Y saa vain kokonaislukuarvoja. Integraalit

$$M^*(t) = \int_0^t \frac{J(s)}{Y(s)}\,\mathrm{d}M_h(s)$$

ovat stokastisia integraaleja lokaalien neliöllisesti integroituvien martingaalien M_h suhteen. Tällöin lemman 2.3 perusteella $M^*(t)$ on lokaali neliöllisesti integroituva martingaali kaikilla t. Nyt $M^*(0) = 0$, jolloin lemman 2.1 nojalla $\mathbb{E}(M^*(t)) = \mathbb{E}(M^*(0)) = 0$. Täten ottamalla odotusarvon lausekkeesta (3.3) saamme

$$\mathbb{E}\Big(\hat{A}_h(t) - A_h^*(t)\Big) = \mathbb{E}\left(\int_0^t \frac{J(s)}{Y(s)}\,\mathrm{d}M_h(s)\right) = 0$$

ja edelleen odotusarvon lineaarisuuden nojalla

$$\mathbb{E}\Big(\hat{A}_h(t)\Big) - \mathbb{E}\Big(A_h^*(t)\Big) = 0.$$

Käyttämällä lauseketta (3.2) saamme Nelson-Aalen estimaattorin $\hat{A}_h(t)$ odotusarvoksi

$$\begin{aligned}
\mathbb{E}\Big(\hat{A}_h(t)\Big) &= \mathbb{E}\Big(A_h^*(t)\Big) \\
&= \mathbb{E}\left(\int_0^t \mu_h^x(s) J(s)\,\mathrm{d}s\right) \\
&= \int_0^t \mu_h^x(s) \mathbb{E}\Big(\mathbb{1}\Big(Y(s) > 0\Big)\Big)\,\mathrm{d}s \\
&= \int_0^t \mu_h^x(s) \mathbb{P}\Big(Y(s) > 0\Big)\,\mathrm{d}s.
\end{aligned}$$

Huomaamme tästä estimaattorin $\hat{A}_h$ olevan harhainen, sillä sen odotusarvo ei ole estimoitava funktio $A_h(t) = \int_0^t \mu_h^x(s)\,\mathrm{d}s$. Voidaan kuitenkin sanoa, että estimaattori on asymptoottisesti harhaton, sillä kuten edellisessä kappaleessa todettiin, vakuutettujen määrän ollessa suuri, todennäköisyys, että ainakin joku selviää vakuutuskauden loppuun on suuri. Tällöin pätee $\mathbb{P}(Y(s) > 0) \to 1$, kun $n \to \infty$. Estimaattorin harha $B(\hat{A}_h(t))$ voidaan laskea ja se on

$$\begin{aligned}
B\Big(\hat{A}_h(t)\Big) &= \mathbb{E}\Big(\hat{A}_h(t)\Big) - A_h(t) \\
&= \int_0^t \mu_h^x(s) \mathbb{P}\Big(Y(s) > 0\Big)\,\mathrm{d}s - \int_0^t \mu_h^x(s)\,\mathrm{d}s \\
&= \int_0^t -\mu_h^x(s)\Big(1 - \mathbb{P}\Big(Y(s) > 0\Big)\Big)\,\mathrm{d}s \\
&= \int_0^t -\mu_h^x(s) \mathbb{P}\Big(Y(s) = 0\Big)\,\mathrm{d}s.
\end{aligned}$$

Kuten jo aiemmin todettiin, otoskoon kasvaessa rajatta $\mathbb{P}(Y(s) = 0) \to 0$, jolloin myös harha suppenee kohti nollaa.

3.2.2 Varianssi

Lähdetään etsimään estimaattorin $\hat{A}_h(t)$ varianssia muodostamalla ensiksi varianssin yhtälö erotukselle $\hat{A}_h(t) - A_h^*(t)$. Varianssin määritelmästä saadaan

$$\mathrm{Var}\Big(\hat{A}_h(t) - A_h^*(t)\Big) = \mathbb{E}\Big(\Big(\hat{A}_h(t) - A_h^*(t)\Big)^2\Big) - \Big(\mathbb{E}\Big(\hat{A}_h(t) - A_h^*(t)\Big)\Big)^2.$$

Edellisessä kappaleessa todettiin, että $\mathbb{E}(\hat{A}_h(t) - A_h^*(t)) = 0$, joten varianssin kaava supistuu muotoon

$$\mathrm{Var}\Big(\hat{A}_h(t) - A_h^*(t)\Big) = \mathbb{E}\Big(\Big(\hat{A}_h(t) - A_h^*(t)\Big)^2\Big).$$

Kuten odotusarvoa laskiessa todettiin $\hat{A}_h(t) - A_h^*(t)$ on lokaali neliöllisesti integroituva martingaali. Huomautuksen 2.11 perusteella prosessin $(\hat{A}_h(t) - A_h^*(t))^2$ kompensaattori on $\langle \hat{A}_h(t) - A_h^*(t) \rangle$, jolloin $(\hat{A}_h(t) - A_h^*(t))^2 - \langle \hat{A}_h(t) - A_h^*(t) \rangle$ on lokaali martingaali, jonka arvo kohdassa 0 on 0. Tällöin varianssi saadaan huomautuksen 2.5 nojalla muotoon

$$\begin{aligned}
\mathrm{Var}\Big(\hat{A}_h(t) - A_h^*(t)\Big) &= \mathbb{E}\langle \hat{A}_h(t) - A_h^*(t) \rangle + \mathbb{E}\Big(\Big(\hat{A}_h(t) - A_h^*(t)\Big)^2 - \langle \hat{A}_h(t) - A_h^*(t) \rangle\Big) \\
&= \mathbb{E}\left\langle \int_0^t \frac{J(s)}{Y(s)} \,\mathrm{d}M_h(s) \right\rangle - 0 \\
&= \mathbb{E}\left\langle \int_0^t \frac{J(s)}{Y(s)} \,\mathrm{d}M_h(s) \right\rangle .
\end{aligned}$$

Käytetään nyt lemman 2.3 tulosta ja ominaisuutta $\langle M_h \rangle = \Lambda_h$, jolloin saadaan

$$\begin{aligned}
\mathbb{E}\left\langle \int_0^t \frac{J(s)}{Y(s)} \,\mathrm{d}M_h(s) \right\rangle &= \mathbb{E}\int_0^t \left(\frac{J(s)}{Y(s)}\right)^2 \,\mathrm{d}\langle M_h(s) \rangle \\
&= \mathbb{E}\int_0^t \left(\frac{J(s)}{Y(s)}\right)^2 \,\mathrm{d}\Lambda_h(s) \\
&= \mathbb{E}\int_0^t \left(\frac{J(s)}{Y(s)}\right)^2 \mu_h^x(s) Y(s) \,\mathrm{d}s \\
&= \mathbb{E}\int_0^t \frac{J(s)}{Y(s)} \mu_h^x(s) \,\mathrm{d}s \\
&= \int_0^t \mathbb{E}\left(\frac{J(s)}{Y(s)}\right) \mu_h^x(s) \,\mathrm{d}s,
\end{aligned}$$

sillä $(J(s))^2 = \mathbb{1}(Y(s) > 0)\mathbb{1}(Y(s) > 0) = \mathbb{1}(Y(s) > 0) = J(s)$. Huomataan, että varianssin lausekkeessa esiintyy kuolevuus μ_h^x, jonka arvoa emme tiedä, joten lähdetään etsimään estimaattoria varianssille. Määritellään estimaattori $\hat{\sigma}_h^2(t) = [\hat{A}_h - A_h^*](t)$, jolloin käyttämällä lemmaa 2.3 ja ominaisuutta $[M_h] = N_h$ saadaan

$$\begin{aligned}
\hat{\sigma}_h^2(t) &= [\hat{A}_h - A_h^*](t) \\
&= \left[\int_0^t \frac{J(s)}{Y(s)} \,\mathrm{d}M_h(s)\right] \\
&= \int_0^t \left(\frac{J(s)}{Y(s)}\right)^2 \,\mathrm{d}[M_h(s)] \\
&= \int_0^t \frac{J(s)}{\big(Y(s)\big)^2} \,\mathrm{d}N_h(s).
\end{aligned}$$

Integraali ei sisällä tuntematonta kuolevuutta, joten sen arvo voidaan jo laskea. Osoitetaan seuraavaksi, että estimaattori $\hat{\sigma}_h^2$ on harhaton estimaattori varianssille. Käyttämällä yhteyttä $\mathrm{d}N_h(s) = \mathrm{d}M_h(s) + \mu_h^x(s)Y(s)\,\mathrm{d}s$ saadaan

$$
\begin{aligned}
B\left(\hat{\sigma}_h^2(t)\right) &= \mathbb{E}\left(\hat{\sigma}_h^2(t)\right) - \mathrm{Var}\Big(\hat{A}_h(t) - A_h^*(t)\Big)\\
&= \mathbb{E}\left(\int_0^t \frac{J(s)}{\big(Y(s)\big)^2}\,\mathrm{d}N_h(s)\right) - \int_0^t \mathbb{E}\left(\frac{J(s)}{Y(s)}\right)\mu_h^x(s)\,\mathrm{d}s\\
&= \mathbb{E}\left(\int_0^t \frac{J(s)}{\big(Y(s)\big)^2}\,\mathrm{d}M_h(s)\right) + \mathbb{E}\left(\int_0^t \frac{J(s)}{\big(Y(s)\big)^2}\mu_h^x(s)Y(s)\,\mathrm{d}s\right)\\
&\quad - \int_0^t \mathbb{E}\left(\frac{J(s)}{Y(s)}\right)\mu_h^x(s)\,\mathrm{d}s\\
&= \mathbb{E}\left(\int_0^t \frac{J(s)}{\big(Y(s)\big)^2}\,\mathrm{d}M_h(s)\right) + \mathbb{E}\left(\int_0^t \frac{J(s)}{Y(s)}\mu_h^x(s)\,\mathrm{d}s\right) - \int_0^t \mathbb{E}\left(\frac{J(s)}{Y(s)}\right)\mu_h^x(s)\,\mathrm{d}s.
\end{aligned}
$$

Ensimmäinen termi on huomautuksen 2.5 nojalla nolla, sillä integraali on lemman 2.3 nojalla lokaali martingaali ja se saa arvon 0 kohdassa 0. Toisessa termissä voidaan puolestaan Fubinin lauseen perusteella vaihtaa odotusarvon ja integraalin paikkaa, sillä integrandi on positiivinen funktio, jolloin harha $B(\hat{\sigma}_h^2(t))$ supistuu muotoon

$$
B\Big(\hat{\sigma}_h^2(t)\Big) = 0 + \int_0^t \mathbb{E}\left(\frac{J(s)}{Y(s)}\mu_h^x(s)\right)\,\mathrm{d}s - \int_0^t \mathbb{E}\left(\frac{J(s)}{Y(s)}\right)\mu_h^x(s)\,\mathrm{d}s = 0.
$$

Edellä muodostettiin yhtälö erotuksen $\hat{A}_h(t) - A_h^*(t)$ varianssille. Haluamme kuitenkin sanoa jotain myös Nelson-Aalen estimaattorin $\hat{A}_h$ varianssista. Oletetaan nyt, että vain pienellä todennäköisyydellä pätee $Y(s) = 0$ jollakin $s \leq t$. Tällöin

$$
\mathbb{E}\Big(\hat{A}_h(t)\Big) = \int_0^t \mu_h^x(s)\mathbb{P}\Big(Y(s) > 0\Big)\,\mathrm{d}s \approx \int_0^t \mu_h^x(s)\,\mathrm{d}s
$$

ja

$$
A_h^*(t) \approx \int_0^t \mu_h^x(s)\,\mathrm{d}s.
$$

Saadaan, että $A_h^*(t) \approx \mathbb{E}(\hat{A}_h(t))$, kun vain pienellä todennäköisyydellä pätee $Y(s) = 0$

jollakin $s \leq t$. Tällöin ainakin suurella otoskoolla pätee

$$\begin{aligned}
\mathrm{Var}\Big(\hat{A}_h(t) - A_h^*(t)\Big) &= \mathbb{E}\Big(\Big(\hat{A}_h(t) - A_h^*(t)\Big)^2\Big) \\
&= \mathbb{E}\Big(\Big(\hat{A}_h(t)\Big)^2 - 2\hat{A}_h(t)A_h^*(t) + \Big(A_h^*(t)\Big)^2\Big) \\
&\approx \mathbb{E}\left(\Big(\hat{A}_h(t)\Big)^2 - 2\hat{A}_h(t)\mathbb{E}\Big(\hat{A}_h(t)\Big) + \Big(\mathbb{E}\Big(\hat{A}_h(t)\Big)\Big)^2\right) \\
&= \mathbb{E}\Big(\Big(\hat{A}_h(t)\Big)^2\Big) - 2\mathbb{E}\Big(\hat{A}_h(t)\Big)\mathbb{E}\Big(\hat{A}_h(t)\Big) + \Big(\mathbb{E}\Big(\hat{A}_h(t)\Big)\Big)^2 \\
&= \mathbb{E}\Big(\Big(\hat{A}_h(t)\Big)^2\Big) - \Big(\mathbb{E}\Big(\hat{A}_h(t)\Big)\Big)^2 \\
&= \mathrm{Var}\Big(\hat{A}_h(t)\Big).
\end{aligned}$$

Tällöin isolla otoskoolla estimaattoria $\hat{\sigma}_h^2$ voidaan käyttää myös Nelson-Aalen estimaattorin $\hat{A}_h$ varianssin estimaattorina.

Aiheesta voi lukea lisää lähteistä Andersen et al. [3] luvusta IV ja Lehtomaa [10].

3.3 Sovellus

Havainnollistetaan tässä luvussa Nelson-Aalen estimaattoria tekemällä lyhyt sovellus R:n avulla. Luomme aluksi tarvittavan datan ja lopuksi piirrämme kuvaajaan Nelson-Aalen estimaatin sekä kumuloivan kuolevuuden todellisen arvon. Tutkimme kuvaajasta vastaako estimaatti datasta laskettua oikeaa arvoa. Todellisuudessa emme tiedä kuolinsyihin liittyvien elinikien jakaumaa, joten emme voi laskea oikeaa arvoa, mutta tehdään se tässä, jotta saadaan jonkinlainen käsitys kuinka hyvin estimaattori vastaa oikeita arvoja. Sovelluksessa käytetty R-koodi löytyy liitteestä A.

Oletetaan, että on olemassa kolme mahdollista eri kuolinsyytä ja oletetaan, että on olemassa $n = 1000$ vakuutettua. Oletetaan esimerkin yksinkertaistamiseksi, että kuolinsyihin liittyvät elinajat noudattavat eksponenttijakaumaa parametreilla $2, 3$ ja 4. Toisin sanoen oletetaan, että

$$\begin{aligned}
T_1(x) &\sim Exp(2), \\
T_2(x) &\sim Exp(3), \\
T_3(x) &\sim Exp(4).
\end{aligned}$$

Tällöin simuloitujen kuolinaikojen arvot eivät vastaa todellisuuden elinikiä, mutta näemme silti kuinka hyvin Nelson-Aalen estimaattien arvot vastaavat todellisia arvoja. Lasketaan ihan aluksi todelliset arvot kumulatiivisille kuolevuuksille. Esimerkin 2.1 nojalla

$\mu_1^x(s) = 2, \mu_2^x(s) = 3$ ja $\mu_3^x(s) = 4$ kaikilla $s \geq 0$, jolloin

$$\int_0^t \mu_1^x(s)\,\mathrm{d}s = 2t,$$
$$\int_0^t \mu_2^x(s)\,\mathrm{d}s = 3t,$$
$$\int_0^t \mu_3^x(s)\,\mathrm{d}s = 4t.$$

Simuloidaan seuraavaksi jokaiselle vakuutetulle i kuhunkin kuolinsyyhyn liittyvä hypoteettinen elinaika $T_h^i(x)$, missä $h = 1, 2, 3$. Saadaan kullekin data $(T_1^i(x), T_2^i(x), T_3^i(x))$ ja kilpailevien kuolinsyiden teorian perusteella vakuutetun i todellinen elinaika on nyt $T^i(x) = \min(T_1^i(x), T_2^i(x), T_3^i(x))$. Todellista elinaikaa vastaava elinaika kertoo kuolinsyyn.

Taulukossa 3.1 nähdään simuloidut kuolinsyihin liittyvät hypoteettiset elinajat kymmenen vakuutetun osalta. Identifioidaan vakuutettu kolumnin ID arvolla.

ID	Kuolinsyy 1	Kuolinsyy 2	Kuolinsyy 3
1	0.532095	0.030062	0.213501
2	2.346608	0.077937	0.326672
3	0.262844	1.213068	0.036988
4	0.562601	0.155092	0.553596
5	0.579605	0.132895	0.20739
6	0.594725	1.788641	0.38414
7	0.218878	0.57054	0.166495
8	0.366237	0.142007	0.856515
9	0.792077	0.111818	0.194791
10	0.001252	0.589447	0.188093

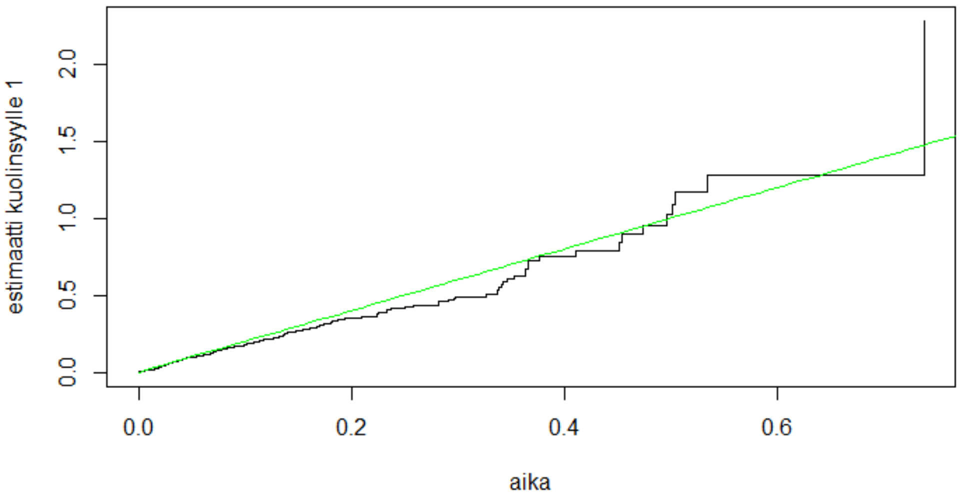
Taulukko 3.1: Simuloidut kuolinsyihin 1, 2 ja 3 liittyvät hypoteettiset elinajat kymmenen vakuutetun osalta.

Seuraavassa taulukossa esitetään todellinen elinaika ja todellista elinaikaa vastaava kuolinsyy kymmenen vakuutetun osalta.

ID	Elinaika	Kuolinsyy
1	0.030062	2
2	0.077937	2
3	0.036988	3
4	0.155092	2
5	0.132895	2
6	0.38414	3
7	0.166495	3
8	0.142007	2
9	0.111818	2
10	0.001252	1

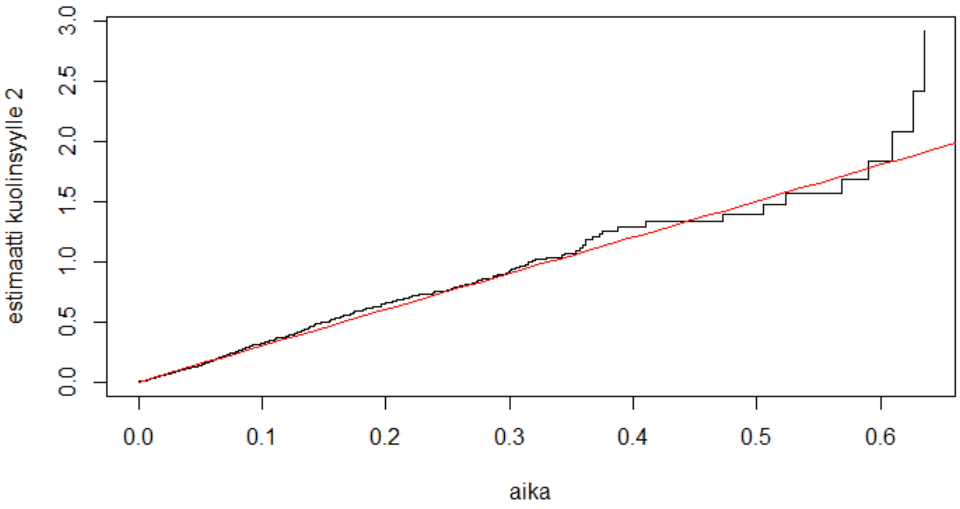
Taulukko 3.2: Havaitut todelliset elinajat ja kuolinsyyt kymmenelle vakuutetulle.

Saadusta datasta voidaan laskea Nelson-Aalen estimaatit. Kuvassa 3.1 on piirrettynä estimaatti kuolinsyylle 1 sekä vihreällä kumulatiivisen kuolevuuden todellinen arvo. Nähdään, että laskettu estimaatti seurailee melko hyvin oikeaa arvoa.

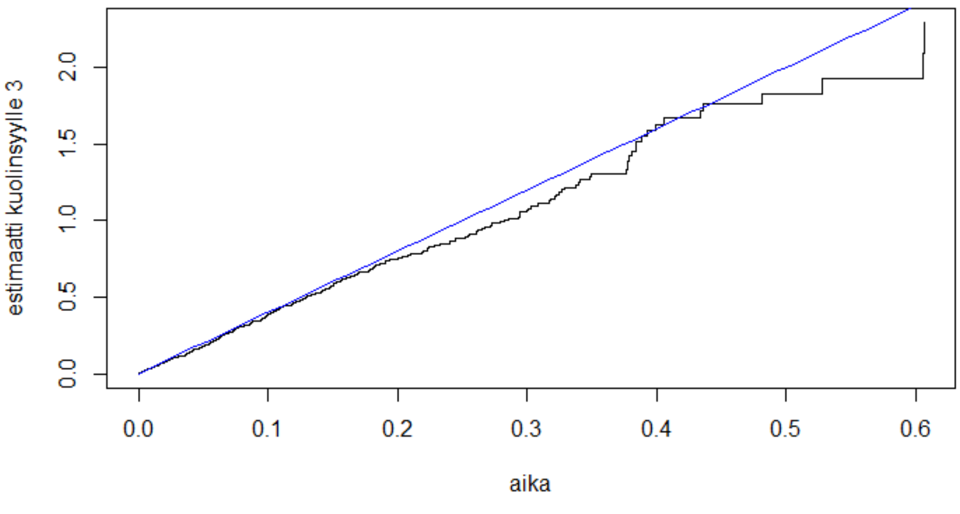


Kuva 3.1: Kuvaajassa mustalla on piirretty estimaatti ja vihreällä todellinen arvo.

Kuvissa 3.2 ja 3.3 nähdään vastaavat estimaatit liittyen kuolinsyihin 2 ja 3.



Kuva 3.2: Kuvaajassa mustalla on piirretty estimaatti ja punaisella todellinen arvo.



Kuva 3.3: Kuvaajassa mustalla on piirretty estimaatti ja sinisellä todellinen arvo.

Kaikissa kuvissa estimaatit näyttävät seurailevan melko hyvin todellista arvoa. Tämä pätee varsinkin aika-akselin alkuosassa, kun havaintoja on paljon. Luonnollisesti havaintojen vähentyessä hajontaa alkaa tulla enemmän. Simulointia toistettaessa saadaan samankaltasia kuvaajia, joissa estimaatti vastaa todellista arvoa varsin hyvin aika-akselin alkupuolella ja havaintojen harventuessa ei enää kovinkaan paljon. Näin ollen Nelson-Aalen menetelmä voi olla hyvä siinä osassa kuolevuusfunktiota, jossa havaintoja on paljon. Toisaalta kysenalaista on millaisia estimaatteja saadaan, jos kuolevuus onkin sellaista muotoa, että aika-akselin alkupuolella havaintoja on vähän ja loppupäässä paljon.

Tässä sovelluksessa on tärkeä huomioida, että emme piirtäneet kuvia estimoiduista kuolevuusfunktioista μ_1^x, μ_2^x ja μ_3^x, vaan estimoiduista kumulatiivisista kuolevuuksista $\int \mu_1^x(u)\,\mathrm{d}u$, $\int \mu_2^x(u)\,\mathrm{d}u$ ja $\int \mu_3^x(u)\,\mathrm{d}u$. Nelson-Aalen estimaattori estimoi juurikin kumulatiivista kuolevuutta. Kumulatiivisesta kuolevuudesta pääseminen itse kuolevuusfunktioon ei ole aivan yksinkertaista, sillä estimaatit eivät ole derivoituvia. Nelson-Aalen estimaattoria on kuitenkin mahdollista tasoittaa, jotta saadaan estimaattori kuolevuusfunktiolle. Aiheesta voi lukea lisää lähteestä Andersen et al. [3] luvusta IV.

4 Lause Nelson-Aalen estimaattorin heikosta konvergenssista

4.1 Konvergenssilauseen määrittely ja todistus

Tutkitaan seuraavaksi Nelson-Aalen estimaattorin ominaisuuksia, kun vakuutettujen määrä n kasvaa rajatta. Merkitään $\hat{A}^{(n)} = (\hat{A}_1^{(n)}, ..., \hat{A}_k^{(n)})$ ja $A = (A_1, ..., A_k)$. Estimaattori $\hat{A}_h^{(n)}$ on tismalleen sama kuin edellisessä luvussa esitelty Nelson-Aalen estimaattori $\hat{A}_h$, mutta korostetaan parametrin n merkinnällä niitä termejä, jotka riippuvat parametrista n, sillä tarkastelemme tilannetta, jossa n kasvaa rajatta. Tutkitaan pelkästään tilannetta, jossa laskuriprosessin $N^{(n)} = (N_1^{(n)}, ..., N_k^{(n)})$ komponenteilla on muotoa $\lambda_h^{(n)}(t) = \mu_h^x(t) Y^{(n)}(t)$ olevat intensiteettiprosessit, joissa μ_h^x on sama kaikilla $n = 1, 2, ...$ Esitellään ja todistetaan seuraavaksi lause, jonka mukaan normalisoitu Nelson-Aalen estimaattori suppenee jakaumaltaan kohti odotusarvoltaan nolla olevaa Gaussista martingaalia. Erityisesti kiinteällä parametrin t arvolla Nelson-Aalen estimaattori on asymptoottisesti normaalijakautunut. Tämä mahdollistaa luottamusvälin määrittämisen prosessille A. Luottamusvälin määrittämisestä on mahdollista lukea lähteestä Andersen et al. [3] luvusta IV.

Määritellään aluksi lausessa käytettävä notaatio.

Määritelmä 4.1. Merkinnällä $D[0, b]^k$ tarkoitetaan avaruutta $\mathbb{R}^k$-arvoisia cadlag funktioita välillä $[0, b]$ varustettuna Skorohod topologialla.

Skorohod topologiasta voi lukea lähteestä Billingsley [5] luvusta 3. Jätetään kuitenkin tässä työssä yksityiskohtien tarkastelu välistä.

Määritelmä 4.2 (Multinormaalijakauma)**.** Satunnaisvektori $U = (U_1, ..., U_k)$ noudattaa multinormaalijakauma $\mathcal{N}(\mu, \Sigma)$, mikäli on mahdollista kirjoittaa

$$U = \mu + AZ,$$

missä $\mu \in \mathbb{R}^k$, A on $k \times l$ -matriisi ja Z noudattaa l-ulotteista standardinormaalijakaumaa. Tällöin $\mathbb{E}(U) = \mu$ ja $\mathrm{Cov}(U) = \Sigma = AA^T$.

Määritelmä 4.3 (Gaussinen prosessi)**.** Gaussinen prosessi on stokastinen prosessi, jolle pätee, että millä tahansa äärellisellä osajoukolla prosessin satunnaismuuttujia on multinormaalijakauma.

Olemme nyt valmiita esittämään konvergenssilauseen.

Lause 4.1 (Nelson-Aalen estimaattorin heikko konvergenssi). *Olkoon $t \in [0, b]$ ja oletetaan, että on olemassa jono positiivisia vakioita $\{a_n\}$, joka kasvaa kohti äärettömyyttä, kun $n \to \infty$. Oletetaan lisäksi, että on olemassa ei-negatiivisia funktiota y_h siten, että μ_h^x / y_h on integroituva välillä $[0, t]$, kun $h = 1, 2, .., k$. Olkoon*

$$\sigma_h^2(s) = \int_0^s \frac{\mu_h^x(u)}{y_h(u)} \, \mathrm{d}u,$$

missä $h = 1, 2, .., k$ ja oletetaan, että

(A) Kaikille $s \in [0, t]$ ja $h = 1, 2, ..., k$ pätee

$$a_n^2 \int_0^s \frac{J^{(n)}(u)}{Y^{(n)}(u)} \mu_h^x(u) \, \mathrm{d}u \xrightarrow{\mathbb{P}} \sigma_h^2(s) \qquad \textit{kun } n \to \infty$$

(B) Kaikille $\epsilon > 0$ ja $h = 1, 2, ..., k$ pätee

$$a_n^2 \int_0^t \frac{J^{(n)}(u)}{Y^{(n)}(u)} \mu_h^x(u) \mathbb{1} \left\{ \left| a_n \frac{J^{(n)}(u)}{Y^{(n)}(u)} \right| > \epsilon \right\} \mathrm{d}u \xrightarrow{\mathbb{P}} 0 \qquad \textit{kun } n \to \infty$$

(C) Kun $h = 1, 2, ..., k$

$$a_n \int_0^t \Big(1 - J^{(n)}(u)\Big) \mu_h^x(u) \, \mathrm{d}u \xrightarrow{\mathbb{P}} 0 \qquad \textit{kun } n \to \infty$$

Tällöin joukossa on $D[0, t]^k$ pätee

$$a_n(\hat{A}^{(n)} - A) \xrightarrow{\mathrm{D}} U = (U_1, ..., U_k) \qquad \textit{kun } n \to \infty,$$

missä $U_1, ..., U_k$ ovat riippumattomia Gaussisia martingaaleja, joille pätee $U_h(0) = 0$ ja $\mathrm{Cov}(U_h(s_1), U_h(s_2)) = \sigma_h^2(\min(s_1, s_2))$.

Huomautus 4.1. On mahdollista edelleen osoittaa, että kiinteällä parametrin t arvolla suurinpiirtein pätee $\hat{A}_h^{(n)}(t) \sim \mathcal{N}(A_h(t), \hat{\sigma}_h^2(t))$, kun n kasvaa rajatta. Tässä $\hat{\sigma}_h^2$ on kappaleessa 3.2.2 esitetty varianssin estimaattori.

Käytetään lauseen todistukseessa lähteen Andersen et al. [3] luvun II versiota Rebolledon keskeisestä raja-arvolauseesta lokaaleille martingaaleille. Luodaan aluksi tarvittava ympäristö raja-arvolauseelle.

Olkoon $n = 1, 2, ...$ Määritellään raja-arvolause tilanteessa, jossa tarkastellaan k-ulotteista laskuriprosessia $N^{(n)} = (N_1^{(n)}, ..., N_k^{(n)})$, jolla on jatkuva kompensaattori. Olkoon laskuriprosessin intensiteettiprosessi $\lambda^{(n)}$. Oletetaan, että $H^{(n)}$ on $k \times k$-ulotteinen matriisi, jonka komponentit ovat lokaalisti rajoitettuja, ennustettavia prosesseja. Olkoon lisäksi $\tilde{M}^{(n)} = (\tilde{M}_1^{(n)}, ..., \tilde{M}_k^{(n)})$ k-ulotteinen vektori, jonka komponentit ovat lokaalisti neliöllisesti integroituvia martingaaleja. Määritellään

$$\tilde{M}_j^{(n)}(t) = \sum_{h=1}^{k} \left(\int_0^t H_{jh}^{(n)}(s) \, \mathrm{d}N_h^{(n)}(s) - \int_0^t H_{jh}^{(n)}(s) \, \lambda_h^{(n)}(s) \mathrm{d}s \right).$$

Seuraamalla lähteen Andersen et al. [3] päättelyä voidaan osoittaa, että tässä tapauksessa

$$\langle \tilde{M}_j^{(n)}, \tilde{M}_{j'}^{(n)} \rangle(t) = \sum_{h=1}^{k} \int_0^t H_{jh}^{(n)}(s) H_{j'h}^{(n)}(s) \, \lambda_h^{(n)}(s) \mathrm{d}s.$$

Jätetään kuitenkin todistus lukijan itsenäisen tarkastelun varaan.

Oletetaan lisäksi, että kaikilla $\epsilon > 0$, $\tilde{M}_\epsilon^{(n)} = (\tilde{M}_{1\epsilon}^{(n)}, ..., \tilde{M}_{k\epsilon}^{(n)})$ on k-ulotteinen vektori, jonka komponentit ovat lokaalisti neliöllisesti integroituvia martingaaleja. Oletetaan, että vektori $\tilde{M}_\epsilon^{(n)}$ sisältää kaikki vektorin $\tilde{M}^{(n)}$ komponenttien hypyt, jotka ovat itseisarvoltaan on suurempia kuin ϵ. Määritellään

$$\begin{aligned} \tilde{M}_{j\epsilon}^{(n)}(t) = \sum_{h=1}^{k} \Bigg(& \int_0^t H_{jh}^{(n)}(s) \mathbb{1}\Big(|H_{jh}^{(n)}(s)| > \epsilon\Big) \, \mathrm{d}N_h^{(n)}(s) \\ & - \int_0^t H_{jh}^{(n)}(s) \mathbb{1}\Big(|H_{jh}^{(n)}(s)| > \epsilon\Big) \, \lambda_h^{(n)}(s) \mathrm{d}s \Bigg). \end{aligned}$$

Jälleen lähteen Andersen et al. [3] perusteella

$$\langle \tilde{M}_{j\epsilon}^{(n)}, \tilde{M}_{j\epsilon}^{(n)} \rangle(t) = \sum_{h=1}^{k} \int_0^t \Big(H_{jh}^{(n)}(s) \Big)^2 \mathbb{1}\Big(|H_{jh}^{(n)}(s)| > \epsilon\Big) \lambda_h^{(n)}(s) \mathrm{d}s.$$

Olkoon $U = (U_1, ..., U_k)$ Gaussinen vektori ja oletetaan, että vektorin U komponentit ovat jatkuvapolkuisia martingaaleja. Oletetaan, että V on jatkuva, deterministinen, positiivisesti semidefiniitti, $k \times k$ matriisiarvoinen funktio välillä $[0, b]$. Oletetaan lisäksi, että funktion V lisäykset ovat positiivisesti semidefiniittejä ja $V(0) = 0$. Oletetaan, että $\langle U \rangle = [U] = V$ ja oletetaan, että $U(0) = 0$. Tällöin $U(t) - U(s)$ noudattaa multinormaalijakaumaa $\mathcal{N}(0, V(t) - V(s))$. Lisäksi kaikilla $0 \leq u \leq s \leq t$ pätee, että $U(t) - U(s)$ on riippumaton vektorista $U(u)$.

Olemme nyt valmiita muotoilemaan lähteen Andersen et al. [3] version Rebolledon keskeisestä raja-arvolauseesta lokaaleille martingaaleille.

Lause 4.2 (Rebolledon keskeinen raja-arvolause lokaaleille martingaaleille)**.** *Oletetaan, että $A \subseteq [0,b]$. Oletetaan, että pätee*

$$\langle \tilde{M}^{(n)} \rangle(t) \xrightarrow{\mathbb{P}} V(t) \qquad \textit{kaikilla } t \in A, \textit{ kun } n \to \infty.$$

Oletetaan lisäksi, että pätee

$$\langle \tilde{M}^{(n)}_{\epsilon h} \rangle(t) \xrightarrow{\mathbb{P}} 0 \qquad \textit{kaikilla } t \in A, h \textit{ ja } \epsilon > 0, \textit{ kun } n \to \infty.$$

Tällöin kaikilla $t_1, ..., t_p \in A$ pätee

$$\Big(\tilde{M}^{(n)}(t_1), ..., \tilde{M}^{(n)}(t_p)\Big) \xrightarrow{\mathrm{D}} \Big(U(t_1), ..., U(t_p)\Big), \qquad \textit{kun } n \to \infty.$$

Jos osajoukko A on välin $[0,b]$ tiheä osajoukko, joka sisältää päätepisteen b, pätee

$$\tilde{M}^{(n)} \xrightarrow{\mathrm{D}} U \qquad \textit{joukossa } D[0,b]^k, \textit{ kun } n \to \infty.$$

Lauseen 4.1 todistus. Todistetaan nyt lause 4.1 käyttämällä Rebolledon keskeistä raja-arvolausetta. Valitaan

$$H^{(n)}_{hj}(s) = \frac{\delta_{hj} a_n J^{(n)}(s)}{Y^{(n)}(s)},$$

missä

$$\delta_{hj} = \begin{cases} 0, & \text{kun } h \neq j \\ 1, & \text{kun } h = j. \end{cases}$$

Nyt $H^{(n)}$ on $k \times k$-matriisi, jonka komponentteja ovat $H^{(n)}_{hj}$, missä $h, j = 1, ..., k$. Matriisin ainoat nollasta poikkeavat arvot löytyvät sen diagonaalilta, jossa kaikki arvot ovat $a_n J^{(n)}/Y^{(n)}$. Toisin sanoen

$$H^{(n)} = \left.\underbrace{\begin{bmatrix} a_n \frac{J^{(n)}}{Y^{(n)}} & 0 & 0 & \cdots & 0 \\ 0 & a_n \frac{J^{(n)}}{Y^{(n)}} & 0 & \cdots & 0 \\ 0 & 0 & a_n \frac{J^{(n)}}{Y^{(n)}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_n \frac{J^{(n)}}{Y^{(n)}} \end{bmatrix}}_{k \text{ saraketta}} \right\} k \text{ riviä.}$$

Prosessit $a_n J^{(n)}/Y^{(n)}$ ovat lokaalisti rajoitettuja ja ennustettavia. Nyt koska intensiteetti on muotoa $\lambda_h^{(n)}(t) = \mu_h^x(t) Y^{(n)}(t)$, saadaan

$$
\begin{aligned}
\tilde{M}_j^{(n)}(t) &= \sum_{h=1}^{k} \Big(\int_0^t H_{jh}^{(n)}(s)\, \mathrm{d}N_h^{(n)}(s) - \int_0^t H_{jh}^{(n)}(s)\, \lambda_h^{(n)}(s) \mathrm{d}s \Big) \\
&= \int_0^t a_n \frac{J^{(n)}(s)}{Y^{(n)}(s)}\, \mathrm{d}N_j^{(n)}(s) - \int_0^t a_n \frac{J^{(n)}(s)}{Y^{(n)}(s)}\, \mu_j^x(s) Y^{(n)}(s) \mathrm{d}s \\
&= a_n \Big(\int_0^t \frac{J^{(n)}(s)}{Y^{(n)}(s)}\, \mathrm{d}N_j^{(n)}(s) - \int_0^t J^{(n)}(s) \mu_j^x(s) \mathrm{d}s \Big) \\
&= a_n \Big(\hat{A}_j^{(n)}(t) - A_j^{*(n)}(t) \Big).
\end{aligned}
$$

Toisaalta kaavan (3.3) nojalla pätee

$$
a_n \Big(\hat{A}_h^{(n)}(t) - A_h^{*(n)}(t) \Big) = a_n \int_0^t \frac{J^{(n)}(s)}{Y^{(n)}(s)}\, \mathrm{d}M_h^{(n)}(s), \qquad \text{missä } h = 1, ..., k.
$$

Täten komponentit $\tilde{M}_j^{(n)}(t)$ ovat lokaaleja neliöllisesti integroituvia martingaaleja. Kun $j = j'$, saadaan

$$
\begin{aligned}
\langle \tilde{M}_j^{(n)}, \tilde{M}_{j'}^{(n)} \rangle(t) &= \sum_{h=1}^{k} \int_0^t H_{jh}^{(n)}(s) H_{j'h}^{(n)}(s)\, \lambda_h^{(n)}(s) \mathrm{d}s \\
&= \int_0^t a_n^2 \frac{J^{(n)}(s)}{\big(Y^{(n)}(s)\big)^2} \mu_j^x(s) Y^{(n)}(s) \mathrm{d}s \\
&= a_n^2 \int_0^t \frac{J^{(n)}(s)}{Y^{(n)}(s)} \mu_j^x(s) \mathrm{d}s.
\end{aligned}
$$

Kun $j \neq j'$, pätee $\langle \tilde{M}_j^{(n)}, \tilde{M}_{j'}^{(n)} \rangle(t) = 0$. Täten $k \times k$-matriisin $\langle \tilde{M}^{(n)} \rangle$ ainoat nollasta poikkeavat arvot ovat sen diagonaalilla.

$$
\langle \tilde{M}^{(n)} \rangle(t) = \left. \underbrace{\begin{bmatrix}
a_n^2 \int_0^t \frac{J^{(n)}(s)}{Y^{(n)}(s)} \mu_1^x(s) \mathrm{d}s & 0 & \cdots & 0 \\
0 & a_n^2 \int_0^t \frac{J^{(n)}(s)}{Y^{(n)}(s)} \mu_2^x(s) \mathrm{d}s & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & a_n^2 \int_0^t \frac{J^{(n)}(s)}{Y^{(n)}(s)} \mu_k^x(s) \mathrm{d}s
\end{bmatrix}}_{k \text{ saraketta}} \right\} k \text{ riviä}
$$

Lauseen 4.1 ehdon (A) mukaan pätee

$$
a_n^2 \int_0^t \frac{J^{(n)}(s)}{Y^{(n)}(s)} \mu_j^x(s) \mathrm{d}s \xrightarrow{\mathbb{P}} \int_0^t \frac{\mu_j^x(s)}{y_j(s)}\, \mathrm{d}s.
$$

Valitaan

$$V(t) = \underbrace{\begin{bmatrix} \int_0^t \frac{\mu_1^x(s)}{y_1(s)}\,\mathrm{d}s & 0 & \cdots & 0 \\ 0 & \int_0^t \frac{\mu_2^x(s)}{y_2(s)}\,\mathrm{d}s & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \int_0^t \frac{\mu_k^x(s)}{y_k(s)}\,\mathrm{d}s \end{bmatrix}}_{k \text{ saraketta}} \left.\vphantom{\begin{bmatrix} \int_0^t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}\right\} k \text{ riviä.}$$

Nyt $k \times k$-matriisiarvoinen funktio $V : [0, b] \to \mathbb{R}^{k \times k}$ on jatkuva ja deterministinen, koska sen kaikki komponentit ovat jatkuvia ja deterministisiä. Olkoon $z = (z_1, ..., z_k)$, missä $z_1, ..., z_k \in \mathbb{R}$, ja merkitään matriisin rivin i ja kolumnin j komponenttia V_{ij}. Tällöin $z^T V z = \sum_{h=1}^k V_{hh} z_k^2 \geq 0$, sillä $V_{ij} \geq 0$ kaikilla $i, j = 1, ..., k$. Toisin sanoen V on positiivisesti semidefiniitti. Lisäksi $V(0) = 0$ ja funktion V lisäykset ovat positiivisesti semidefiniittejä.

Nyt

$$\begin{aligned} \langle \tilde{M}_{j\epsilon}^{(n)}, \tilde{M}_{j\epsilon}^{(n)} \rangle(t) &= \sum_{h=1}^k \int_0^t \Big(H_{jh}^{(n)}(s) \Big)^2 \mathbb{1}\Big(|H_{jh}^{(n)}(s)| > \epsilon \Big) \lambda_h^{(n)}(s) \mathrm{d}s \\ &= \int_0^t a_n^2 \frac{J^{(n)}(s)}{\Big(Y^{(n)}(s)\Big)^2} \mathbb{1}\left(\left| a_n \frac{J^{(n)}(s)}{Y^{(n)}(s)}(s) \right| > \epsilon \right) \mu_j^x(s) Y^{(n)}(s) \mathrm{d}s \\ &= a_n^2 \int_0^t \frac{J^{(n)}(s)}{Y^{(n)}(s)} \mathbb{1}\left(\left| a_n \frac{J^{(n)}(s)}{Y^{(n)}(s)}(s) \right| > \epsilon \right) \mu_j^x(s) \mathrm{d}s, \end{aligned}$$

joka suppenee lauseen 4.1 ehdon (B) mukaan stokastisesti kohti nollaa, kun n kasvaa rajatta. Täten Rebolledon raja-arvolauseen nojalla pätee $a_n(\hat{A}^{(n)} - A^{*(n)}) \xrightarrow{\mathrm{D}} U$, kun $n \to \infty$.

Koska termi $\left| a_n \Big(A_j^{*(n)}(s) - A_j(s) \Big) \right|$ voidaan kirjoittaa muodossa

$$\begin{aligned} \left| a_n \Big(A_j^{*(n)}(s) - A_j(s) \Big) \right| &= \left| a_n \Big(\int_0^s J^{(n)}(u) \mu_j^x(u) \mathrm{d}u - \int_0^s \mu_j^x(u) \mathrm{d}u \Big) \right| \\ &= \left| a_n \int_0^s \mu_j^x(u) \Big(J^{(n)}(u) - 1 \Big) \mathrm{d}u \right| \\ &= a_n \int_0^s \mu_j^x(u) \Big(1 - J^{(n)}(u) \Big) \mathrm{d}u, \end{aligned}$$

voidaan ehdon (C) perusteella päätellä, että

$$\sup_{s \in [0,t]} \left| a_n \Big(A_h^{*(n)}(s) - A_h(s) \Big) \right| \xrightarrow{\mathbb{P}} 0, \qquad \text{kun } n \to \infty.$$

Proposition 2.1 kohdan 2 nojalla stokastisesta suppenemisesta seuraa jakaumasuppeneminen, joten lauseen 4.1 väite

$$a_n(\hat{A}^{(n)} - A) \xrightarrow{\mathrm{D}} U = (U_1, ..., U_k) \qquad \text{kun } n \to \infty$$

seuraa. Ehto $U_h(0) = 0$ pätee ja kovarianssiehto puolestaan saadaan seuraavasti. Jos $s_1 < s_2$, niin

$$\begin{aligned}\mathrm{Cov}\Big(U_h(s_1), U_h(s_2)\Big) &= \mathrm{Cov}\Big(U_h(s_1), U_h(s_2) - U_h(s_1) + U_h(s_1)\Big)\\ &= \mathrm{Cov}\Big(U_h(s_1), U_h(s_2) - U_h(s_1)\Big) + \mathrm{Cov}\Big(U_h(s_1), U_h(s_1)\Big)\\ &= \mathrm{Var}\Big(U_h(s_1)\Big),\end{aligned}$$

sillä riippumattomuuden nojalla $\mathrm{Cov}(U_h(s_1), U_h(s_2) - U_h(s_1)) = 0$. Vastaavasti, jos pätee $s_1 > s_2$, saadaan

$$\begin{aligned}\mathrm{Cov}\Big(U_h(s_1), U_h(s_2)\Big) &= \mathrm{Cov}\Big(U_h(s_1) - U_h(s_2) + U_h(s_2), U_h(s_2)\Big)\\ &= \mathrm{Cov}\Big(U_h(s_1) - U_h(s_2), U_h(s_2)\Big) + \mathrm{Cov}\Big(U_h(s_2), U_h(s_2)\Big)\\ &= \mathrm{Var}\Big(U_h(s_2)\Big).\end{aligned}$$

Täten $\mathrm{Cov}(U_h(s_1), U_h(s_2)) = \mathrm{Var}(U_h(\min(s_1, s_2)))$. Raja-arvolauseen ehtojen perusteella $U(t) - U(s)$ noudattaa multinormaalijakaumaa $\mathcal{N}(0, V(t) - V(s))$. Tällöin $U(t) - U(0)$ noudattaa jakaumaa $\mathcal{N}(0, V(t) - V(0))$, missä $U(0) = 0$ ja $V(0) = 0$. Täten

$$\mathrm{Cov}\Big(U_h(s_1), U_h(s_2)\Big) = \int_0^{\min(s_1,s_2)} \frac{\mu_h^x(u)}{y_h(u)} \,\mathrm{d}u = \sigma_h^2\Big(\min(s_1, s_2)\Big).$$

□

4.2 Konvergenssilauseen sovellus kilpailevien kuolinsyiden tilanteeseen

Sovelletaan lausetta 4.1 luvussa 3 kuvailtuun tilanteeseen seuraamalla lähteen Andersen et al. [3] luvun IV päättelyä. Yksinkertaistetaan tilannetta hieman ja tarkastellaan konvergenssilausetta välin $[0, b]$ osavälillä $[t_1, t_2]$, missä $P_0(0, s) > 0$ kaikilla $s \in [t_1, t_2]$. Esitellään aluksi muutama tulos, jotka auttavat lauseen ehtojen tarkastamisessa ja näytetään sitten, että lause 4.1 pätee luvun 3 kilpailevien kuolinsyiden tilanteessa.

Esitetään aluksi lähteen Andersen et al. [3] versio Hellandin (1983) konvergenssituloksesta.

Propositio 4.1 (Helland 1983). *Oletetaan, että jollekin jonolle prosesseja $X^{(n)}$ pätee*

$$X^{(n)}(s) \xrightarrow{\mathbb{P}} f(s) \qquad \textit{kun } n \to \infty$$

melkein kaikilla $s \in [0, b]$, missä f on deterministinen funktio. Oletetaan lisäksi, että $X^{(n)}(s)$, missä $n = 1, 2, 3, ...$, on tasaisesti integroituva eli

$$\lim_{C\to\infty} \sup_n \mathbb{E}\Big(\Big|X^{(n)}(s)\Big|\mathbb{1}(\Big|X^{(n)}(s)\Big| > C)\Big) = 0 \qquad \textit{kaikilla } s.$$

Oletetaan, että kaikille s, n pätee

$$\mathbb{E}\Big|X^{(n)}(s)\Big| \leq k(s),$$

missä

$$\int_0^b k(s)\,\mathrm{d}s < \infty.$$

Tällöin funktiolle f pätee

$$\int_0^b |f(s)|\,\mathrm{d}s < \infty$$

sekä

$$\mathbb{E}\left(\sup_t \Big|\int_0^t X^{(n)}(s)\,\mathrm{d}s - \int_0^t f(s)\,\mathrm{d}s\Big|\right) \to 0 \qquad \textit{kun } n \to \infty.$$

Esitellään seuraavaksi Alan Gutin kirjan *Probabability: a Graduate Course* [8] tulos, joka tarjoaa vaihtoehtoisen tavan tasaisesti integroituvuuden osoittamiseen.

Lemma 4.1. *Olkoon $X^{(1)}, X^{(2)}, ...$ satunnaismuuttujia. Oletetaan, että*

$$\sup_n \mathbb{E}(|X^{(n)}|^p) < \infty \qquad \textit{jollain } p > 1.$$

Tällöin $\{X^{(n)}, n \geq 1\}$ on tasaisesti integroituva.

Lemman 4.1 todistus. Oletetaan, että $p > 1$. Voidaan kirjoittaa

$$|X^{(n)}|^p \mathbb{1}(|X^{(n)}| > a) = |X^{(n)}|\,|X^{(n)}|^{p-1}\mathbb{1}(|X^{(n)}| > a) \geq |X^{(n)}|\,a^{p-1}\mathbb{1}(|X^{(n)}| > a),$$

joten saadaan epäyhtälö $|X^{(n)}|^p\mathbb{1}(|X^{(n)}| > a) \geq |X^{(n)}|\,a^{p-1}\mathbb{1}(|X^{(n)}| > a)$, jota pyörittelemällä saadaan

$$|X^{(n)}|\mathbb{1}(|X^{(n)}| > a) \leq a^{1-p}\,|X^{(n)}|^p\mathbb{1}(|X^{(n)}| > a).$$

Käyttämällä saatua epäyhtälöä saadaan

$$\begin{aligned}\mathbb{E}\Big(|X^{(n)}|\mathbb{1}(|X^{(n)}| > a)\Big) &\leq a^{1-p}\,\mathbb{E}\Big(|X^{(n)}|^p\mathbb{1}(|X^{(n)}| > a)\Big)\\ &\leq a^{1-p}\,\mathbb{E}(|X^{(n)}|^p)\\ &\leq a^{1-p}\sup_n \mathbb{E}(|X^{(n)}|^p) \quad \to 0, \quad \text{kun } a \to \infty,\end{aligned}$$

joka pätee riippumatta parametrista n. Lisäksi

$$\begin{aligned}\mathbb{E}\Big(|X^{(n)}|\mathbb{1}(|X^{(n)}| > a)\Big) &\geq a\mathbb{P}(|X^{(n)}| > a)\\ &\geq a \cdot 0\\ &= 0.\end{aligned}$$

Täten X_n on tasaisesti integroituva. □

Muotoillaan seuraavaksi Ian W McKeague, Klaus J Utikal, et al. artikkelissa *Inference for a nonlinear counting process regression model* [12] esitetty Lemma 1. Kyseessä on tekninen tulos, joka auttaa ehtojen (A)-(C) tarkastamisessa.

Lemma 4.2. *Olkoon X binomijakautunut satunnaismuuttuja parametrilla (n, p), missä $0 < p \leq 1$. Määritellään $1/X = 0$, jos $X = 0$.*

Tällöin pätee

$$\mathbb{E}\left(\frac{1}{X}\right)^k \leq \left(\frac{k+1}{np}\right)^k \qquad \textit{kaikilla } k = 1, 2, 3, ...$$

Lemman 4.2 todistus. Oletetaan, että X on binomijakautunut satunnaismuuttuja parametrilla (n, p), missä $0 < p \leq 1$. Merkitään $q = 1 - p$ ja oletetaan, että k on positiivinen kokonaisluku. Tällöin

$$\begin{aligned}\mathbb{E}\left(\frac{1}{X}\right)^k &= \sum_{i=1}^n \frac{1}{i^k}\,\mathbb{P}(X = i)\\ &= \sum_{i=1}^n \frac{1}{i^k}\frac{n!}{i!(n-i)!}p^i q^{n-i}\\ &= \sum_{i=1}^n \frac{1}{i^k}\frac{(i+k)!}{i!(i+k)!}\frac{n!}{(n-i)!}p^i q^{n-i}\\ &= \sum_{i=1}^n \frac{1}{i^k}\frac{(i+1)...(i+k)}{(i+k)!}\frac{n!}{(n-i)!}p^i q^{n-i}\\ &= \sum_{i=1}^n \frac{i+1}{i}\frac{i+2}{i}...\frac{i+k}{i}\frac{1}{(i+k)!}\frac{n!}{(n-i)!}p^i q^{n-i}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \left(1+\frac{1}{i}\right)\left(1+\frac{2}{i}\right)...\left(1+\frac{k}{i}\right)\frac{1}{(i+k)!}\frac{n!}{(n-i)!}p^i q^{n-i} \\
&\leq \sum_{i=1}^{n} (k+1)^k \frac{n!}{(i+k)!(n-i)!}p^i q^{n-i} \\
&= \frac{(k+1)^k n!}{p^k (n+k)!}\sum_{i=1}^{n} \frac{(n+k)!}{(i+k)!(n-i)!}p^{i+k} q^{n-i} \\
&= \frac{(k+1)^k}{p^k}\frac{1}{(n+1)...(n+k)} \cdot 1 \\
&\leq \frac{(k+1)^k}{p^k}\frac{1}{\underbrace{n...n}_{\text{k kpl}}} \\
&= \left(\frac{k+1}{np}\right)^k.
\end{aligned}
$$

□

Tarkistetaan ensiksi (A) kohdan ehto. Tavoitteena on tarkastaa proposition 4.1 oletukset ja osoittaa ehdon (A) pätevyys proposition tuloksen avulla.

Tiedämme, että kun otoskoko kasvaa rajatta, $J^{(n)}(s) \to 1$. Tiedämme lisäksi, että $Y^{(n)}(s) = \sum_{i=1}^{n} \mathbb{1}(T^i(x) \geq s)$, missä vakuutettujen elinajat ovat toisistaan riippumattomia ja samoin jakautuneita. Tällöin otoskoon n kasvaessa rajatta vahvan suurten lukujen lain nojalla $Y^{(n)}(s)/n \xrightarrow{m.v.} \mathbb{E}(\mathbb{1}(T^i(x) \geq s))$. Nyt $\mathbb{E}(\mathbb{1}(T^i(x) \geq s)) = \mathbb{P}(T^i(x) \geq s)) = P_0(0,s)$, jolloin $Y^{(n)}(s)/n \xrightarrow{m.v.} P_0(0,s)$, kun $n \to \infty$. Proposition 2.1 kohdan 1 nojalla melkein varmasta suppenemisesta seuraa suppeneminen stokastisessa mielessä, joten voimme kirjoittaa

$$
n\frac{J^{(n)}(s)}{Y^{(n)}(s)}\mu_h^x(s) \xrightarrow{\mathbb{P}} \frac{\mu_h^x(s)}{P_0(0,s)} \qquad \text{kun } n \to \infty,
$$

kaikilla $s \in [t_1, t_2]$. Termi $Y^{(n)}(s)$ on binomijakautunut parametrillä $(n, P_0(0,s))$, sillä voimme ajatella termin $Y^{(n)}(s)$ laskemista toistokokeena. Toisin sanoen tarkastellaan jokaisen vakuutetun kohdalla ovatko he tilassa 0 juuri ennen hetkeä s. Huomataan, että pätee $0 < P_0(0,s) \leq 1$.

Kertomalla puolittain parametrillä n, voimme muokata lemman 4.2 epäyhtälön muotoon

$$
\mathbb{E}\left(\frac{n}{X}\right)^k \leq \left(\frac{k+1}{p}\right)^k.
$$

Lemman oletuksissa määriteltiin, että $1/X = 0$, jos $X = 0$, joten saadaan

$$\mathbb{E}\left(\frac{n}{X}\right)^k = \mathbb{E}\left(\frac{n\mathbb{1}(X>0) + n\mathbb{1}(X=0)}{X}\right)^k = \mathbb{E}\left(\frac{n\mathbb{1}(X>0)}{X} + 0\right)^k,$$

jolloin pätee

$$\mathbb{E}\left(\frac{n\mathbb{1}(X>0)}{X}\right)^k \leq \left(\frac{k+1}{p}\right)^k.$$

Epäyhtälö pätee erityisesti silloin, kun $k = 1$.

Olkoon $s \in [t_1, t_2]$ kiinnitetty. Kuten yllä totesimme $Y^{(n)}(s)$ on binomijakautunut, joten lemman 4.2 nojalla epäyhtälö

$$\mathbb{E}\left(\frac{n\mathbb{1}\left(Y^{(n)}(s) > 0\right)}{Y^{(n)}(s)}\right) \leq \frac{2}{P_0(0,s)}$$

pätee. Kertomalla epäyhtälön molemmat puolet kuolevuudella $\mu_h^x(s)$ saadaan

$$\mathbb{E}\left(\frac{nJ^{(n)}(s)\mu_h^x(s)}{Y^{(n)}(s)}\right) \leq \frac{2\mu_h^x(s)}{P_0(0,s)},$$

missä

$$\int_{t_1}^{t_2} \frac{2\mu_h^x(s)}{P_0(0,s)} \, \mathrm{d}s < \infty,$$

sillä oletimme, että $\int_0^t \mu_h^x(s) \, \mathrm{d}s < \infty$ kaikilla h ja kaikilla $t \in [0, b]$.

Osoitetaan seuraavaksi, että $\{nJ^{(n)}(s)\mu_h^x(s)(Y^{(n)}(s))^{-1} : n = 1, 2, ...\}$ on tasaisesti integroituva. Jälleen käyttämällä hyväksi lemmaa 4.2 saadaan

$$\mathbb{E}\left(\left(\frac{nJ^{(n)}(s)\mu_h^x(s)}{Y^{(n)}(s)}\right)^2\right) \leq \left(\frac{3\mu_h^x(s)}{P_0(0,s)}\right)^2 < \infty.$$

Siis $\mathbb{E}((nJ^{(n)}(s)\mu_h^x(s)(Y^{(n)}(s))^{-1})^2)$ on rajoitettu ja koska raja ei riipu parametrista n voidaan puhua tasaisesti rajoitettuvuudesta parametrin n suhteen. Täten käyttämällä lemmaa 4.1 voidaan osoittaa, että $\{nJ^{(n)}(s)\mu_h^x(s)(Y^{(n)}(s))^{-1} : n = 1, 2, ...\}$ on tasaisesti integroituva kaikilla $s \in [t_1, t_2]$.

Nyt, kun valitaan

$$X^{(n)}(s) = n\frac{J^{(n)}(s)}{Y^{(n)}(s)}\mu_h^x(s), \qquad f(s) = \frac{\mu_h^x(s)}{P_0(0,s)} \qquad \text{ja} \qquad k(s) = \frac{2\mu_h^x(s)}{P_0(0,s)},$$

pätee proposition 4.1 nojalla

$$\mathbb{E}\left(\sup_s\left|\int_{t_1}^s X^{(n)}(u)\,\mathrm{d}u - \int_{t_1}^s f(u)\,\mathrm{d}u\right|\right) \to 0 \qquad \text{kun } n \to \infty.$$

Tästä seuraa edelleen

$$\int_{t_1}^s n\frac{J^{(n)}(u)}{Y^{(n)}(u)}\mu_h^x(u)\,\mathrm{d}u \xrightarrow{\mathbb{P}} \int_{t_1}^s \frac{\mu_h^x(u)}{P_0(0,u)}\,\mathrm{d}u \qquad \text{kaikilla } s \in [t_1, t_2],$$

sillä stokastinen suppeneminen seuraa L^1 suppenemisesta. Täten olemme osoittaneet, että lauseen 4.1 ehto (A) pätee, kun valitaan $a_n = \sqrt{n}$ ja $y_h(u) = P_0(0,u)$ kaikilla h, jolloin $\sigma_h^2(s) = \int_{t_1}^s \frac{\mu_h^x(u)}{P_0(0,u)}\,\mathrm{d}u$.

Kohta (B) voidaan tarkastaa vastaavasti. Nyt pätee

$$\sqrt{n}\frac{J^{(n)}(s)}{Y^{(n)}(s)} = \frac{1}{\sqrt{n}}\frac{nJ^{(n)}(s)}{Y^{(n)}(s)}.$$

(A) kohdan tarkastelun nojalla $\frac{nJ^{(n)}(s)}{Y^{(n)}(s)}$ suppenee stokastisesti kohti $\frac{1}{P_0(0,s)}$, kun n kasvaa rajatta, mutta koska kertoimena on $\frac{1}{\sqrt{n}}$, pätee

$$\sqrt{n}\frac{J^{(n)}(s)}{Y^{(n)}(s)} \xrightarrow{\mathbb{P}} 0, \qquad \text{kun } n \to \infty$$

kaikilla $s \in [t_1, t_2]$. Tästä seuraa, että kaikilla $\epsilon > 0$ ja $s \in [t_1, t_2]$

$$n\frac{J^{(n)}(s)}{Y^{(n)}(s)}\mu_h^x(s)\mathbb{1}\left\{\left|\sqrt{n}\frac{J^{(n)}(s)}{Y^{(n)}(s)}\right| > \epsilon\right\} \xrightarrow{\mathbb{P}} 0, \qquad \text{kun } n \to \infty.$$

Lisäksi (A) kohdan tarkistuksen nojalla pätee

$$\mathbb{E}\left(n\frac{J^{(n)}(s)}{Y^{(n)}(s)}\mu_h^x(s)\mathbb{1}\left\{\left|\sqrt{n}\frac{J^{(n)}(s)}{Y^{(n)}(s)}\right| > \epsilon\right\}\right) \leq \mathbb{E}\left(\frac{nJ^{(n)}(s)\mu_h^x(s)}{Y^{(n)}(s)}\right) \leq \frac{2\mu_h^x(s)}{P_0(0,s)}$$

sekä

$$\begin{aligned}\mathbb{E}\left(\left(n\frac{J^{(n)}(s)}{Y^{(n)}(s)}\mu_h^x(s)\mathbb{1}\left\{\left|\sqrt{n}\frac{J^{(n)}(s)}{Y^{(n)}(s)}\right| > \epsilon\right\}\right)^2\right) &\leq \mathbb{E}\left(\left(\frac{nJ^{(n)}(s)\mu_h^x(s)}{Y^{(n)}(s)}\right)^2\right) \\ &\leq \left(\frac{3\mu_h^x(s)}{P_0(0,s)}\right)^2 < \infty.\end{aligned}$$

Täten ehdon (B) tarkistus etenee kuin (A) kohdassa.

Tarkistetaan lopuksi ehto (C). Koska μ_h^x on rajoitettu välillä $[0,t]$ kaikilla $t \in [0,b]$, voidaan määritellä

$$\hat{\mu}_h = \max\{\mu_h^x(s) \,:\, s \in [t_1, t_2]\}.$$

Merkitään tämän lisäksi hetkeä, jolloin viimeinen vakuutettu kuolee T^*, toisin sanoen

$$T^* = \max\Big(T^1(x), ..., T^n(x)\Big).$$

Tällöin voidaan kirjoittaa

$$\begin{aligned}
\sqrt{n}\int_{t_1}^{t_2}\Big(1 - J^{(n)}(u)\Big)\mu_h^x(u)\,\mathrm{d}u &\leq \sqrt{n}\int_{t_1}^{t_2}\Big(1 - J^{(n)}(u)\Big)\hat{\mu}_h\,\mathrm{d}u \\
&= \sqrt{n}\hat{\mu}_h\left(\int_{t_1}^{t_2} 1\,\mathrm{d}u - \int_{t_1}^{t_2}\mathbb{1}\Big(Y^{(n)}(u) > 0\Big)\,\mathrm{d}u\right) \\
&= \sqrt{n}\hat{\mu}_h\left((t_2 - t_1) - \int_{t_1}^{t_2}\mathbb{1}(T^* \geq u)\,\mathrm{d}u\right) \\
&= \sqrt{n}\hat{\mu}_h\left((t_2 - t_1) - \int_{t_1}^{\min(t_2, T^*)} 1\,\mathrm{d}u\right) \\
&= \sqrt{n}\hat{\mu}_h\Big((t_2 - t_1) - \min(t_2, T^*) + t_1\Big) \\
&= \sqrt{n}\hat{\mu}_h\Big(t_2 - \min(t_2, T^*)\Big) \\
&= \sqrt{n}\hat{\mu}_h(t_2 - T^*)\mathbb{1}(T^* < t_2) \\
&\leq \sqrt{n}\hat{\mu}_h t_2 \mathbb{1}(T^* < t_2).
\end{aligned}$$

Todistetaan seuraavaksi, että $\sqrt{n}\hat{\mu}_h t_2 \mathbb{1}(T^* < t_2)$ suppenee L^1 mielessä kohti nollaa. Vakuutettujen eliniät ovat riippumattomia, samoin jakautuneita satunnaismuuttujia, joten $\mathbb{P}(T^* < t_2) = (\mathbb{P}(T^1(x) < t_2))^n$. Täten voidaan kirjoittaa

$$\begin{aligned}
\mathbb{E}\Big(|\sqrt{n}\hat{\mu}_h t_2 \mathbb{1}(T^* < t_2) - 0|\Big) &= \mathbb{E}\Big(\sqrt{n}\hat{\mu}_h t_2 \mathbb{1}(T^* < t_2)\Big) \\
&= \sqrt{n}\hat{\mu}_h t_2 \mathbb{P}(T^* < t_2) \\
&= \sqrt{n}\hat{\mu}_h t_2 \Big(\mathbb{P}(T^1(x) < t_2)\Big)^n \\
&= \sqrt{n}\hat{\mu}_h t_2 \exp\left(n \log\Big(\mathbb{P}(T^1(x) < t_2)\Big)\right).
\end{aligned}$$

Nyt $\log(\mathbb{P}(T^1(x) < t_2)) < 0$, joten kun annetaan vakuutettujen määrän n kasvaa rajatta $\mathbb{E}(|\sqrt{n}\hat{\mu}_h t_2 \mathbb{1}(T^* \leq t_2) - 0|) \to 0$. Proposition 2.1 kohdan 3 perusteella L^1 suppenemisesta seuraa stokastinen suppeneminen, joten

$$\sqrt{n}\int_{t_1}^{t_2}\Big(1 - J^{(n)}(u)\Big)\mu_h^x(u)\,\mathrm{d}u \leq \sqrt{n}\hat{\mu}_h t_2 \mathbb{1}(T^* \leq t_2) \xrightarrow{\mathbb{P}} 0, \qquad \text{kun } n \to \infty.$$

Lisäksi

$$\sqrt{n}\int_{t_1}^{t_2}\Big(1-J^{(n)}(u)\Big)\mu_h^x(u)\,\mathrm{d}u \geq 0 \xrightarrow{\mathbb{P}} 0, \qquad \text{kun } n\to\infty.$$

Täten saadaan

$$\sqrt{n}\int_{t_1}^{t_2}\Big(1-J^{(n)}(u)\Big)\mu_h^x(u)\,\mathrm{d}u \xrightarrow{\mathbb{P}} 0, \qquad \text{kun } n\to\infty.$$

Nyt lauseen 4.1 jokaisen ehdon on todistettu pätevän, joten saadaan, että välillä $[t_1, t_2]$ pätee

$$(\sqrt{n}(\hat{A}_h^{(n)} - A_h) : h \neq 0) \xrightarrow{\mathrm{D}} (U_h : h \neq 0),$$

missä U_h ovat riippumattomia, Gaussisia martingaaleja, joille pätee

$$U_h(0) = 0 \qquad \text{ja} \qquad \mathrm{Cov}\Big(U_h(s), U_h(u)\Big) = \sigma_h^2(s\wedge u) = \int_{t_1}^{s\wedge u}\frac{\mu_h^x(v)}{P_0(0,v)}\,\mathrm{d}v.$$

5 Muita tapoja ratkaista estimointiongelma

Nelson-Aalen estimaattorin lisäksi on olemassa myös muita tapoja estimoida kuolevuutta. Esitellään tässä luvussa niistä muutama lyhyesti.

5.1 Parametrisoidut mallit

Kuolevuusfunktiota voidaan estimoida yksinkertaisemmin funktioilla. Alla on mainittu näistä pari tunnetuinta.

Gompertzin kuolevuusfunktio esitetään muodossa

$$\mu(x) = be^{cx},$$

missä $b, c > 0$ ovat vakioita. Vakiolla b kuvataan lähtökohtaista kuolevuutta syntymähetkellä ja c puolestaan kuvaa suhdetta tai tahtia, jolla kuolevuus kasvaa iän myötä.

Makeham laajensi Gompertzin mallia lisäämällä siihen vakion $a > 0$

$$\mu(x) = a + be^{cx}.$$

Parametrin a voidaan ajatella kuvaavan onnettomuuksiin liittyvää kuolevuutta.

Weibullin malli on puolestaan muotoa

$$\mu(x) = bx^d,$$

missä $b, d > 0$ ovat vakioita.

Lisää malleista voi lukea luentomonisteesta Nyrhinen [14] luvusta 3 sekä Missov et al. [13]

5.2 Kuolevuustaulu

Kuolevuustaulu (*engl. life table, mortality table*) on eräs yleisimmistä tavoista arvioida kuolevuutta.

Kohortti kohtaisessa kuolevuustaulu -mallissa ajatus on seuraava. Oletetaan, että l_0 ihmistä syntyy samanaikaisesti. Seurataan kaikkia kuolemaan asti ja kerätään jokaiselta eliniältä $x = 0, 1, 2, ...$ tietoa, kuinka moni kuoli välillä x ja $x + 1$. Kerätään myös tietoa, kuinka moni selviytyi ikään (syntymäpäivään) x asti. Merkitään

$$l_x = \text{ ikään } x \text{ selvinneiden yksilöiden lukumäärä}$$

ja

$$d_x = l_x - l_{x+1} = \text{ ikävälillä } x \text{ ja } x + 1 \text{ kuolleiden lukumäärä.}$$

Kun laajennetaan x saamaan arvoja reaaliluvuilta, määritellään selviytymisfunktio

$$S(x) = \frac{l_x}{l_0}.$$

Tällöin voidaan laskea, kuinka suuri osa alkuperäisestä joukosta (kohortista) kuolee aikavälillä x ja $x+t$

$$\frac{S(x) - S(x+t)}{S(x)} = \frac{l_x - l_{x+t}}{l_x}.$$

Laskut suorittamalla nähdään, millä aikavälillä on suurin riski kuolla. Nyt

$$q_x = \frac{d_x}{l_x} = \text{ todennäköisyys kuolla aikavälillä } x \text{ ja } x+1$$

ja

$$p_x = 1 - q_x.$$

Ikä x	l_x	d_x	x	l_x	d_x
0	100000	2629	40	92315	295
1	97371	141	41	92020	332
2	97230	107	42	91688	408
3	97123	63	43	91280	414
4	97060	63	44	90866	464
5	96997	69	45	90402	532
6	96928	69	46	89870	587
7	96859	52	47	89283	680
8	96807	54	48	88603	702
9	96753	51	49	87901	782
10	96702	33	50	87119	841
11	96669	40	51	86278	885
12	96629	47	52	85393	974
13	96582	61	53	84419	1082
14	96521	86	54	83337	1088
15	96435	105	55	82249	1213
16	96330	83	56	81036	1344
17	96247	125	57	79692	1423
18	96122	133	58	78269	1476
19	95989	149	59	76793	1572
20	95840	154	60	75221	1696
21	95686	138	61	73525	1784
22	95548	163	62	71741	1933
23	95385	168	63	69808	2022
24	95217	166	64	67786	2186
25	95051	151	65	65600	2261
26	94900	149	66	63339	2371
27	94751	166	67	60968	2426
28	94585	157	68	58542	2356
29	94428	133	69	56186	2702
30	94295	160	70	53484	2548
31	94135	149	71	50936	2677
32	93986	152	72	48259	2811
33	93834	160	73	45448	2763
34	93674	199	74	42685	2710
35	93475	187	75	39975	2848
36	93288	212	76	37127	2832
37	93076	228	77	34295	2835
38	92848	272	78	31460	2803
39	92576	261			

Taulukko 5.1: Lähteessä Slud [16] simuloitu kohortti-kuolevuustaulu, joka pyrkii mallintamaan yhdysvaltalaisten miesten kohortti-kuolevuustaulua ikään 78 asti.

5.3 Referenssikuolevuudet k2004 ja k2012

Referenssikuolevuudella tarkoitetaan suomalaisten henkivakuutusyhtiöiden vakuuttamien suomalaisten toteutunutta ja toteutuvaksi ennustettua kuolevuutta. Henkivakuutuksen referenssikuolevuuden k2004 pohjana on Mika Mäkisen SHV-työ *Referenssikuolevuus henkivakuutusyhtiöille* (2004). Työssä on käytetty HMD-tietokannasta (Human Mortality Database) löytyvää Suomen populaatioaineistoa ja henkivakuutusyhtiöiltä kerättyä riskiperusteaineistoa, josta on saatu henkivakuutusaineiston havaittu kuolevuus vuosina 1972-2001.

Mäkinen on aluksi tutkinut Suomen väestökuolevuusaineistossa kuolevuuden kehitystä ja tekemällä sen perusteella ennusteet kehityksestä tulevaisuudessa Lee-Carter menetelmällä. Sen jälkeen on arvoitu populaatiokuolevuuden ja henkivakuutuskuolevuuden suhteellista eroa kaavalla

$$p = \frac{\mu_x^{\text{'henkiaineisto'}}}{\mu_x^{\text{'populaatio'}}}.$$

Tämän suhteen on oletettu pysyvän suurin piirtein vakiona. Tästä edelleen on johdettu ennusteisiin sopiva funktiomuotoinen referenssikuolevuus, joka ottaa parametreinaan vakuutetun sukupuolen sp, iän x sekä syntymävuoden sv. Merkitään miestä $sp = 0$ ja naista $sp = 1$. Referenssikuolevuus k2004 on muotoa

$$\mu_s p(x, sv) = \max\Big(0.0001, a_{sp}(sv)\Big) + e^{c_{sp}(x,sv)},$$

missä

$$\begin{aligned} a_0(sv) &= 10^{-5} \cdot 0.744 \cdot (2070 - sv), \\ a_1(sv) &= 10^{-5} \cdot 0.206 \cdot (2019 - sv) \end{aligned}$$

ja

$$\begin{aligned} c_0(x, sv) = 0.05438 \cdot (1716 - sv) &+ 0.000533 \cdot (sv - 1719) \cdot x \\ &- 0.000217 \cdot (sv - 1843) \cdot \max(0, x - 81) \end{aligned}$$

sekä

$$\begin{aligned} c_0(x, sv) = -11.51 &+ 0.000316 \cdot (2253 - sv) \cdot x \\ &+ 0.000783 \cdot (sv - 1916) \cdot \max(0, x - 71). \end{aligned}$$

Aiheesta on mahdollista lukea lisää SHV-tutkinnon Vakuutusmatematiikan sovellukset -materiaalista [1].

Referenssikuolevuuteen on sittemmin tehty päivitys k2012. Henkivakuutusyhtiöiden kuolevuustutkimuksessa k2012 kuolevuusennuste on tehty vakuutuslajeittain, mitä ei oltu tehty k2004 tutkimuksessa. Lisäksi tutkimus k2012 käyttää Lee-Carter menetelmän sijaan Lee-Miller menetelmää. Tutkimuksesta voi lukea tarkemmin lähteestä Sirén et al. [15].

5.4 Kaplan-Meier

Usein kumulatiivisen kuolevuuden estimointi ei riitä, vaan halutaan estimoida selvitymisfunktiota S_0. Lauseen 2.1 nojalla arvioidaan funktiota

$$S(t) = S_0(t) = \exp\left(-\int_0^t \mu(s)\,\mathrm{d}s\right).$$

Merkitään laskuriprosessin N hyppyhetkiä $s_1 < s_2 < \dots$. Tällöin Kaplan-Meier estimaattori on muotoa

$$\hat{S}(t) = \prod_{s_j \le t} \left(1 - \frac{1}{Y(s_j)}\right). \tag{5.1}$$

Intuitiivisesti Kaplan-Meier estimaattori voidaan perustella jakamalla aikaväli $[0, t]$ pieniin väleihin $0 = t_0 < t_1 < \dots < t_K = t$ ja kirjoittamalla selviytymisfunktio S muodossa

$$S(t) = \mathbb{P}(T > t) = \prod_{k=1}^{K} \mathbb{P}(T > t_k \,|\, T > t_{k-1}).$$

Oletetaan, että aikavälit ovat niin tiiviitä, että vain yksi laskuriprosessin hyppy voi tapahtua kullakin välillä. Jos vakuutettu ei kuole aikavälillä $(t_{k-1}, t_k]$, estimoidaan todennäköisyyttä $\mathbb{P}(T > t_k \,|\, T > t_{k-1})$ arvolla 1. Jos puolestaan vakuutettu kuolee aikavälillä $(t_{k-1}, t_k]$, estimoidaan todennäköisyyttä $\mathbb{P}(T > t_k \,|\, T > t_{k-1})$ arvolla $1 - 1/Y(t_{k-1}) = 1 - 1/Y(s_j)$. Täten saadaan kaavan (5.1) muoto estimaattorille.

Lisää aiheesta löytyy Odd Aalenin, Ornulf Borganin ja Hakon Gjessingin kirjasta *Survival and event history analysis: a process point of view* [2].

6 Johtopäätökset

Estimaattorin vahvuudet vaikuttavat tulevan esiin vakuutettujen määrän kasvaessa rajatta. Kuten aiemmin todettiin heikon konvergenssilauseen yhteydessä $\sqrt{n}(\hat{A} - A)$ alkaa muistuttaa jakaumaltaan Gaussista martingaalia, kun n kasvaa rajatta. Tästä edelleen voisimme osoittaa, että otoskoon kasvaessa rajatta kiinteällä parametrin t arvolla suurinpiirtein pätee

$$\hat{A}_h(t) \sim \mathcal{N}(A_h(t), \hat{\sigma}_h^2(t)),$$

missä $\hat{\sigma}_h^2(t)$ on kappaleessa 3.2.2 esitetty estimaattori varianssille. Lähteen Andersen et al. [3] luvun IV mukaan tuloksen avulla saadaan muodostettua luottamusrajat kumulatiiviselle kuolevuudelle A.

Nelson-Aalen estimaattorin asymptoottinen harhattomuus tukee estimaattorin käyttöä suuressa vakuutuskannassa. Lisäksi etua on, että mallissa emme oleta kuolevuudelle tiettyä parametrista muotoa.

Mallin heikkouksia voivat olla muun muassa tietyt oletukset, joita tehtiin matkan varrella. Esimerkiksi oletimme, että kuolinsyy on yksikäsitteisesti määriteltävissä. Oletus ei välttämättä ole täysin realistinen todellisuudessa. Vakuutettujen elinikien riippumattomuuden olettaminenkaan ei aina ole täysin realistista. Näiden lisäksi menetelmä vaatii suuren aineiston ja estimaaatti voi olla myös haastavampi laskea kuin parametrisoidun mallin estimaatti, sillä tavoitteena on ratkaista koko funktio muutaman parametrin sijaan. Kappaleessa 3.3 tehdyn sovelluksen perusteella on lisäksi mahdollista, että Nelson-Aalen menetelmä sopii vain siihen osaan kuolevuusfunktiota, josta on paljon havaintoja.

Yhteenvetona: pienessä vakuutuskannassa Nelson-Aalen estimaattorin laskeminen ei oletettavasti ole järkevää, päinvastoin kuin suuressa vakuutuskannassa, jossa estimaattori voidaan laskea vaikka diagnostiikan vuoksi.

Kirjallisuutta

[1] Shv-tutkintovaatimukset - vakuutusmatematiikan sovellukset 2019- materiaali.

[2] Odd Aalen, Ornulf Borgan, and Hakon Gjessing. *Survival and event history analysis: a process point of view.* Springer Science & Business Media, 2008.

[3] Per Kragh Andersen, Ø rnulf Borgan, Richard D. Gill, and Niels Keiding. *Statistical models based on counting processes.* Springer Series in Statistics. Springer-Verlag, New York, 1993.

[4] Mathias Beiglboeck, Walter Schachermayer, and Bezirgen Veliyev. A short proof of the doob–meyer theorem. *Stochastic Processes and their Applications*, 122(4):1204–1209, 2012.

[5] Patrick Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, 1968.

[6] Patrick Billingsley. *Probability and measure.* John Wiley & Sons, 2008.

[7] David CM Dickson, Mary Hardy, Mary R Hardy, and Howard R Waters. *Actuarial mathematics for life contingent risks.* Cambridge University Press, 2013.

[8] Allan Gut. *Probability: a graduate course*, volume 75. Springer Science & Business Media, 2013.

[9] Konstantin Izyurov. Probability theory -luentomoniste, syksy 2018.

[10] Jaakko Lehtomaa. Advanced life insurance mathematics spring 2019 -luentomoniste.

[11] Jaakko Lehtomaa. Life insurance mathematics 2 spring 2019 -luentomoniste.

[12] Ian W McKeague, Klaus J Utikal, et al. Inference for a nonlinear counting process regression model. *The Annals of Statistics*, 18(3):1172–1187, 1990.

[13] Trifon I Missov, Adam Lenart, Laszlo Nemeth, Vladimir Canudas-Romo, and James W Vaupel. The gompertz force of mortality in terms of the modal age at death. *Demographic Research*, 32:1031–1048, 2015.

[14] Harri Nyrhinen. Life insurance mathematics 1 fall 2018 -luentomoniste.

[15] Tarja Sirén et al. Henkivakuutusyhtiöiden kuolevuustutkimus k2012. 2012.

[16] Eric V Slud. *Actuarial mathematics and life-table statistics.* Chapman & Hall/CRC, 2012.

A Liitteet

```
#Simulointi

#Kuolinsyiden lukumäärä = 3

#Vakuutettujen lukumäärä
n<-1000

set.seed(372529)

#Kuolinsyihin liittyvät hypoteettiset elinajat
data_1<-rexp(n,(2))
data_2<-rexp(n,(3))
data_3<-rexp(n,(4))

id <-seq(1:n)

data_temp<-c(id,data_1,data_2,data_3)

#Luodaan taulukko, jossa on vakuutettujen id:t
#ja eri kuolinsyihin liittyvät eliniat
data<-as.data.frame(
  matrix(data_temp,
         nrow = n,
         ncol = 4,
         byrow = FALSE))
colnames(data)<-
  c("id","kuolinsyy 1","kuolinsyy 2","kuolinsyy 3")

#Luodaan taulukko, jossa on kolumnit
#id ja todellinen elinaika
ajat<-as.data.frame(apply(data[,2:4], 1, FUN=min))
colnames(ajat)<-c("elinaika")
ajat$id<-seq(1:n)

#Luodaan taulukko, jossa on kolumnit id ja kuolinsyy
syyt<-as.data.frame(apply( data[,2:4], 1, which.min))
colnames(syyt)<-c("kuolinsyy")
```

```r
syyt$id<-seq(1:n)

#Luodaan taulukko, jossa on yhdistetty kaikki tiedot.
#Kolumnit ovat id, todelliset elinajat
#ja kuolinsyy
tulos <- merge(ajat,syyt,by="id")

#Järjestetään taulukko kuolinajan mukaan
j_tulos <- tulos[order(tulos$elinaika),]

#Varmistus, että kaikki kuolemat ovat omilla ajanhetkilla
d<-j_tulos[,2]
d_unique<-unique(d)
length(d_unique)

#Luodaan kolumni, jota käytetään laskennassa
laskenta <- c(seq(nrow(j_tulos)))
j_tulos$laskenta<-laskenta

#Luodaan kolumni elossa, joka laskee
#kuinka monta vakuutettua on ollut elossa juuri
#ennen kuolinhetkeä ja luodaan kolumni, joka laskee
#1/(vakuutettujen määrä, jotka ovat olleet elossa
#juuri ennen kuolinhetkeä)
j_tulos$elossa <- with(j_tulos,n-laskenta+1)
j_tulos$suhteellinen <- with(j_tulos, 1/(n-laskenta+1))


#Luodaan taulukko, jossa on vain
#syyhyn 1 kuolleet vakuutetut
tulos_syy_1 = j_tulos[j_tulos$kuolinsyy ==1,]

#Luodaan kolumni, joka laskee kumulatiivista
#summaa kolumnin suhteellinen arvoista
tulos_syy_1$kumulatiivinen <-
  cumsum(tulos_syy_1$suhteellinen)

#Lisätään piirtämistä varten nollarivi
tulos_syy_1 <- rbind(tulos_syy_1,"temp"= c(0,0,1,0,0,0,0))
```

```
tulos_syy_1 <-
  tulos_syy_1[order(tulos_syy_1$laskenta,decreasing=FALSE),]

plot(kumulatiivinen~elinaika,
     data=tulos_syy_1[tulos_syy_1$kuolinsyy==1,],
     type="s",
     xlim=c(0,max(tulos_syy_1$elinaika)),
     ylim=c(0,max(tulos_syy_1$kumulatiivinen)),
     xlab="aika",ylab="estimaatti␣kuolinsyylle␣1")
curve(2*x,from=0,to=1.1, col="green",add=TRUE)


#Luodaan taulukko, jossa on vain
#syyhyn 2 kuolleet vakuutetut
tulos_syy_2 = j_tulos[j_tulos$kuolinsyy ==2,]
tulos_syy_2$kumulatiivinen <-
  cumsum(tulos_syy_2$suhteellinen)

tulos_syy_2 <- rbind(tulos_syy_2,"temp"= c(0,0,1,0,0,0,0))
tulos_syy_2 <-
  tulos_syy_2[order(tulos_syy_2$laskenta,decreasing=FALSE),]

plot(kumulatiivinen~elinaika,
     data=tulos_syy_2[tulos_syy_2$kuolinsyy==2,],
     type="s",
     xlim=c(0,max(tulos_syy_2$elinaika)),
     ylim=c(0,max(tulos_syy_2$kumulatiivinen)),
     xlab="aika",ylab="estimaatti␣kuolinsyylle␣2")
curve(3*x,from=0,to=1.1, col="red",add=TRUE)


#Luodaan taulukko, jossa on vain
#syyhyn 3 kuolleet vakuutetut
tulos_syy_3 = j_tulos[j_tulos$kuolinsyy ==3,]
tulos_syy_3$kumulatiivinen <-
  cumsum(tulos_syy_3$suhteellinen)

tulos_syy_3 <- rbind(tulos_syy_3,"temp"= c(0,0,1,0,0,0,0))
tulos_syy_3 <-
```

```
  tulos_syy_3[order(tulos_syy_3$laskenta,decreasing=FALSE),]

plot(kumulatiivinen~elinaika,
     data=tulos_syy_3[tulos_syy_3$kuolinsyy==3,],
     type="s",
     xlim=c(0,max(tulos_syy_3$elinaika)),
     ylim=c(0,max(tulos_syy_3$kumulatiivinen)),
     xlab="aika",ylab="estimaatti␣kuolinsyylle␣3")
curve(4*x,from=0,to=1.1, col="blue",add=TRUE)
```